National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

NL-339 (r. 88/04) c

Canada

THE UNIVERSITY OF ALBERTA

PATTERNS IN CHAOS, THE NONLINEAR GEODYNAMICS OF EARTHQUAKE PREDICTION

by

Qing Li

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR IF PHILOSOPHY

IN

GEOPHYSICS

DEPARTMENT OF PHYSICS

EDMONTON, ALBERTA

Spring 1991

National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-66731-1

Canada

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR Qing Li

TITLE OF THESIS PATTERNS IN CHAOS, THE NONLINEAR GEODYNAMICS OF EARTHQUAKE PREDICTION

DEGREE FOR WHICH THESIS WAS PRESENTED Doctor of Philosophy

YEAR THIS DEGREE GRANTED Spring 1991

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

(SIGNED) .....

PERMANENT ADDRESS:

9805 - 88 Ave

Edmonton, Alberta, Canada

T6E 2R2

DATED Dec 4, 1990

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled PATTERNS IN CHAOS; THE NONLINEAR GEODYNAMICS OF EARTHQUAKE PREDICTON submitted by QING LI in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY in GEOPHYSICS

.....
Supervisor: Dr. Edo Nyland

.....
Dr. E. R. Kanasewich

.....
Dr. W. Israel

.....
Dr. G. Rostoker

.....
Dr. D. Scott

External Examiner:
Academician Vladimir Isackovich Keilis-Borok
International Institute of Earthquake Prediction
Theory and Mathematical Geophysics
Academy of Sciences of the USSR, Moscow

Date December 4, 1990

This thesis is dedicated to my parents and my wife Jing

in gratitude.

Abstract

The dynamics of the lithosphere is a non-linear deterministic system, described by yet unknown differential equations. The phase portrait of such a system can be reconstructed from a time series of a single observable, an earthquake catalog, which I interpret as a collection of Poincare points in the phase space. Using this point of view the system can be analyzed by the methods of chaos which have recently become important for the study of non-linear systems. The low dimensionality of the attractor, between 3 and 4, obtained in the analysis of seismicity data for the west coast of Canada supports the argument that the dynamics of the lithosphere is chaotic rather than random. The earthquake catalog therefore contains all the information, in the sense of topological equivalence, about the dynamics of the lithosphere; this information can be retrieved for earthquake prediction.

This justifies a pattern recognition study of seismicity on the West coast of Canada with a view to predicting earthquakes of magnitude 6.4 or greater. The results of this approach using the Russian M8 algorithm are promising. The premonitory seismic patterns in the earthquake catalog for the west coast of Canada bear strong similarity to those found in other areas of the world. This implies that they are intrinsic to the dynamics of the lithosphere, and do not depend on local geological structure and tectonic regime. This similarity suggests the existence of an underlying order in lithosphere dynamics.

When catalogs of earthquakes are analysed from this point of view it becomes clear that aftershocks contain more information about integral dynamics of the lithosphere than was previously suspected. Removal of aftershocks from earthquake catalog could result in increased noise in the catalog. This suggests that aftershocks have more significance for earthquake prediction and as evidence of processes in lithosphere dynamics than currently assumed.

The rate of loss of information imposes an upper limit of predictability of the system, and it can be estimated from the catalog. Such an analysis shows that the upper limit of the time for earthquake prediction for the west coast of Canada is about 20 years.

Such limit is intrinsic to the dynamics of the lithosphere in that area.

The chaotic behavior of the lithosphere is probably caused by interaction of fault hierarchies and the non-linear constitution of fault zone. This can be demonstrated in a simple thermally activated instability model but such a model has only limited value in the analysis of hierarchical fault systems in real lithosphere. A general model describing the dynamics of such a fractured lithosphere can be approached by introducing two state variables. These reflect, to a certain extent, the complexity and interaction in sub-fault system. The general calculations required in this theory are extensive but a one dimensional model illustrates the approach and contains a surprising amount of physics.

Acknowledgements

I would like to express my gratitude to my supervisor Dr. Edo Nyland for his patience, guidance, and gentle prods throughout the course of this study. To him I owe many thanks for his encouragement of my independence in research, and for his interactions that constantly bring me back to the right track of my study. I have benefitted a lot from him on how to think as a physicist, and that has had a great influence on my outlook as a scientist.

I owe many thanks to prof. V. I. Keilis-Borok. He left *chaos* behind after his visit to the department, and that inspired this study from the very beginning. Thank you for your advice and understanding for the short period of time in the summer of 1987 when I worked with you on detecting premonitory seismic patterns near the Vancouver Island area.

I take this opportunity to thank Dr. A. Gabrielov and Dr. N. Nazer for sharing your valuable ideas and many useful discussions while working on the shear instability model of seismicity in the spring of 1988, and also for the many good times we spent together in the work.

I would like to thank Dr. D.H. Weichert for his work on the earthquake catalog of the west coast of Canada.

Without the generous support of the Department of Physics at University of Alberta in the form of teaching and researching assistantships this work would not have been possible.

I would like to thank my parents who have wondered what I have been doing for all these years on the other side of the Pacific Ocean, but have never failed to give me less than their complete support.

Last but not least, I would like to thank my wife Jing for her unconditional support in many ways, and for her patience to bear a husband who created more and more chaos at home during this long period of study. I thank my son "Tutu" for making this one of the most enjoyable periods of my life.

TABLE OF CONTENTS

CHAPTER PAGE

INTRODUCTION 1

1 The problems of the lithosphere dynamics 2

2 Earthquake catalogs and their conventional interpretation 4

3 Aftershocks and Omori's law 6

4 Self-similarity and magnitude-frequency relation 7

5 Some earthquake related physical models 10

6 Lithosphere and its dynamical evolution 12

7 General ideas of premonitory seismic patterns 14

8 Tectonic regime and seismicity of the Vancouver Island area 16

CHAPTER 1: A SHEAR INSTABILITY MODEL OF SEISMICITY..... 22

1.1 Thermally activated instability: quasistatic model 22

1.2 A single degree of freedom autonomous system 26

1.3 The effects of the real fault gouge 32

1.4 The effects of inhomogeneity of a fault system 34

1.5 The model of the subduction zone near Vancouver Island, Canada..... 37

CHAPTER 2: IS THE DYNAMICS OF THE LITHOSPHERE CHAOTIC 54

2.1 Chaos theory and its implications on lithosphere dynamics 54

2.2 Reconstruction of the lithosphere dynamics from earthquake catalog..... 58

2.3 Measuring chaos in lithosphere dynamics 61
2.4 Application of singular spectrum analysis to earthquake catalog 67
2.5 Discussion 68

CHAPTER 3: THE LIMIT ON EARTHQUAKE PREDICTION 78
3.1 Information theory and dynamics 79
3.2 Minimum information distribution and the earthquake cycle 81
3.3 Rate of loss of stored information 83
3.4 Estimate the upper limit of time on earthquake prediction 86

CHAPTER 4: DETECTING PREMONITORY SEISMIC PATTERNS 96
4.1 General idea of pattern recognition 97
4.2 Definition of the functions of algorithm M8 99
4.3 Diagnosing the premonitory seismic patterns near Vancouver Island, Canada 101
4.4 Discussion 106

CHAPTER 5: ON MODELLING THE LITHOSPHERE DYNAMICS 122
5.1 Mesoscopic structure of the lithosphere 124
5.2 Phenomenological formulation in mesodomain 126
5.3 Equation of state for a one dimensional material 131
5.4 Discussion 133

CONCLUDING REMARKS 140

BIBLIOGRAPHY 142

APPENDIX A 149

APPENDIX B 153

APPENDIX C. 155

List of Tables

Table		Page
4.1	The time and space windows for removing aftershocks	117
4.2	A two dimensional time magnitude histogram of events in the catalog.	118
4.3	The nine strongest earthquakes in the study region.	119
4.4	A summary of the TIPs for the Vancouver Island area.	120
4.5	Magnitude-time histograms of events in regions centered on two large earthquakes	121

List of Figures

Figure		Page
1	An illustration of slip-weakening model (a) and velocity-weakening model (b).	18
2	The idea that the lithosphere displays the hierarchical structure. Fracturing occurs on all scales, and has a fractal geometry of fault distribution.	19
3	The tectonic map of the west coast of Canada. It shows the main lithosphere plate boundaries and the relative plate motion	20
4	The geographic distribution of epicenters for events of magnitude 4 or greater on the west coast of Canada.	21
1.1	Illustration of the idea that the thermally activated shear instability as a mechanism of earthquakes.	39
1.2	Illustration of a one-dimensional model and its equivalent form described by a spring and dash pot.	40
1.3	The clarification of the phase portraits at the neighborhood of the fixed point.	41
1.4	An example of Hopf bifurcation.	42
1.5	The relation between the real part of eigenvalue of the Jacobian and the loading rate.	43
1.6	The relation between the real part of eigenvalue of the Jacobian and the ambient temperature.	44
1.7	A limit cycle solution of the system.	45
1.8	The stress versus time for the one fault segment model.	46
1.9	Illustration of the idea that the effect of normal pressure is taken into account by introducing the factor f, which is the ratio between the real contact area and apparent contact area.	47
1.10	The relation between the real part of eigenvalue of the Jacobian	

and the factor *f*. 48

1.11 The stress versus time for the model of two fault segments. 49

1.12 The one-dimensional map of successive maximum stresses. 50

1.13 The finite element model of the west coast of Canada. 51

1.14 The magnitudes of simulated earthquakes versus time. 52

1.15 The magnitude and frequency relation of the simulated earthquakes. 53

2.1 Illustration of baker's transform. 71

2.2 Illustration of the idea that earthquake catalogs correspond to a set of Poincare points. 72

2.3 Definition of fractal dimension. 73

2.4 A logC(ε) versus log ε plot for different D calculated for the mainshock catalog which contains events of magnitude above 4. 74

2.5 A logC(ε) versus log ε plot for different D calculated for the mainshock catalog which contains events of magnitude above 2. 75

2.6 A logC(ε) versus log ε plot for different D calculated for the catalog that contains all the events of magnitude above 2. 76

2.7 The results of SSA. 77

3.1 (a) The view of a dynamic system as a communication channel. (b) Partition of a continuous variable. 90

3.2 Illustration of the minimum information distribution for a harmonic oscillator. 91

3.3 Illustration of phase flow of a dynamic system corresponds to K=0, K>0 and K=∞. 92

3.4 General behavior of information I(t) that decreases with time. 93

3.5 The result of calculation of entropy K_2 for the catalog containing all events of magnitude above 3. 94

3.6 The result of calculation of entropy K_2 for the catalog containing

	all events of magnitude above 2.	95
4.1	General idea of pattern recognition.	108
4.2	An example of one dimensional distribution.	109
4.3	The idea of integral description for diagnosis of TIP.	110
4.4	The largest earthquakes in 1^0x1^0 rectangles.	111
4.5	The centers of the regions analyzed with M8 and the results.	112
4.6	The regions covered by overlapping TIPs and one time intervals common to all TIPs for each event.	113
4.7	The results of the analysis using only events whose depths have been determined to be 25 km or less.	114
4.8	The voting of each function in establishing the TIPs in the analysis	115
4.9	Plots of activation and quiescence for the four large events in the catalog	116
5.1	Illustration of the idea of dividing the lithosphere into three phases according to three measuring scales.	135
5.2	The description of the position, rotation and deformation of a block in the block system.	136
5.3	A one dimensional block model.	137
5.4	The constitutive relation of one dimensional model.	138
5.5	Sketches of a two dimensional model. Fault is described as a weak zone where blocks are randomly organized, and corresponds a state with high entropy.	139
A.1	The definition of a bilinear element and the shear strain.	152

Introduction

There are innumerable problems in geophysics that concern the dynamics of the lithosphere and primary among those is the place of earthquakes in its dynamical development. The dynamics of the lithosphere can be completely described in terms of classical physics but approximation has to be used for mathematical tractability in any particular case. Most attempts use linear models to explore the consequences of particular aspects of the dynamical behavior of the lithosphere and ignore its non-linearity. For example, linear elasticity, or visco-elasticity will give a good approximation to wave propagation in the lithosphere. The stress distribution analysis in the lithosphere by linear elasticity will show an average effect of the stress states due to small deformation. Beyond this approximation, many problems related to the dynamics of the lithosphere require consideration of the non-linearity of the processes. In those circumstances, linear theory could be not only ineffective but even misleading.

The approximation of linear theory is obviously insufficient to describe the state of the lithosphere when an earthquake is about to occur. Under these circumstances, the dynamical behavior of a fault system becomes more than a simple sum of each individual aspect of the fault behavior. It must include all the effects together because it is necessary to treat the dynamics of the lithosphere as a whole in a search for the gross integrated characteristics of lithosphere dynamics. It is the purpose of this dissertation to explore this non-linear dynamical behavior of the lithosphere in the light of recent progress in non-linear dynamics; both in the context of the earthquake prediction in particular, and in the context of the dynamics of the lithosphere in general. I hope this study will provide a better understanding of those premonitory seismicity patterns precede strong earthquakes and justify the pattern recognition technique for earthquake prediction on the basis of theory of nonlinear dynamics.

In this introduction I review the phenomena and recent analyses that suggest that theories of non-linear dynamics may be useful in the study of lithosphere dynamics. The review is certainly not complete and is not intended to be. It presents those notions that appear to me to be central to the problem of developing a viable theory of lithosphere

dynamics.

1 The Problems of lithosphere dynamics

The most apparent manifestation of non-linear lithosphere dynamics is the occurrence of strong earthquakes and their prediction has both scientific and practical interest. In the pursuit of this objective, lack of a theory for the non-linear dynamics of the lithosphere led Keilis-Borok and his colleagues to seek an empirical approach to the problem of earthquake prediction (Keilis-Borok, et al., 1988). They use a pattern recognition technique to identify premonitory seismicity patterns that may precede a strong earthquake and their results are encouraging; 80% of strong earthquakes can be predicted within 1 to 2 years and a few hundred kilometers (Keilis-Borok, 1990). World wide tests of the algorithm show that the earthquake behavior is similar in different regions, independent of tectonic environment and level of seismicity.

Despite its success, the limit of this approach are not clear. For instance, there may be a limit to the amount of information useful for predicting future earthquakes that can be retrieved from even a perfect earthquake catalog. Answers to such questions do not come from the algorithm itself but from the fundamental principles of lithosphere dynamics. Such questions could be explored by modelling but the existence of widely distributed fault segments with complex local geology and tectonic regime raise tremendous obstacles to mathematical modelling of their dynamical behavior. As a result little progress has occurred in the study of the dynamics of the lithosphere in the last few decades.

At the preliminary stage of research on fault physics and mechanics, a simple model that deals with the essential phenomenology is fundamental. In chapter 1 I construct such a model of a single fault along a subducting oceanic plate for the Vancouver Island region on the west coast of Canada. The fault includes effects of temperature and displays high non-linearity. Although the model is simple, it preserves some of the qualitative and quantitative behavior we see in a real fault. The thermal instability in this model is one tempting phenomenon. An artificial earthquake catalog will be generated by taking the occurrence of such an instability as an earthquake. Clearly this model have little predictive

power, but it may have value as a mathematical analog of a fault system.

Unlike the situation for simplified models, the differential equations for the lithosphere and therefore their solutions are unknown. In this case we cannot study the system in a traditional way. But recent developments of chaos theory provide an alternative; it allows us to proceed directly to qualitative information about the system by applying a geometric analysis method. In chapter 2 and chapter 3, I discuss such a geometrical approach to the dynamics of the lithosphere. Here I view the lithosphere dynamic system as a black box which produces a stream of outputs without considering the detailed fault structures or the physical properties of each fault. The best record of such outputs is seismicity, and all our questions about the dynamics of the lithosphere are posed therefore in terms of earthquake catalogs. The catalogs will be used to reconstruct the phase portrait of lithosphere dynamics, and thereafter I will determine some geometric and dynamical invariants, such as dimensionality and Kolmogorov entropy of an underlying strange attractor of the system. These invariants can be used to quantify the chaos or characterize the randomness of the system. In particular I discuss whether the dynamics of the lithosphere is a stochastic process or a deterministic chaos. Understanding this question will help us model the dynamics of the lithosphere.

In chapter 4, I connect these ideas to observations by introducing the basic idea of pattern recognition algorithm on earthquake prediction proposed Keilis-Borok and his colleagues. The M8 algorithm can diagnose premonitory seismic patterns for the Vancouver Island region that are similar to those detected in other areas of the world. I discuss those patterns in terms of the dynamics of the lithosphere, and link them to the geometric patterns on the attractor. In this way we can use our knowledge about the attractor in the dynamics of the lithosphere to impose limits on the pattern recognition technique and to guide further improvement of the algorithm.

Although the lithosphere appears to be a low dimensional non-linear dynamic system, what is responsible for this non-linearity and what kind of mathematical model can describe the dynamics is still not clear. This makes analysis of the dynamics of the lithosphere a more difficult topic than that of other non-linear dynamic systems. In chapter 5 I argue that the non-linearity in the dynamics of the lithosphere results from the non-linear

constitutive relation among stress, strain and other state variables and the development of the state of a faulted region. In addition to stress and strain, I introduce two more state variables to describe the mechanical state of the lithosphere. The dynamical relation of those state variables develops from the methods of statistical physics and they reflect the complexity and interactions in a system.

2 Earthquake Catalogs and Their Conventional Interpretation

Seismicity data is central to the theoretical developments presented here and some discussion of its nature will be useful to non specialists. In this section I introduce some useful jargon and point interested readers to more thorough discussions.

An earthquake catalog is a collection of earthquake events. Each event usually has five components: occurrence time, longitude, latitude, depth and magnitude. The magnitude of an earthquake is an estimate of the elastic energy released during the fracture process. C. F. Richter first suggested a logarithmic scale could characterize the local earthquake magnitude M_L in terms of the maximum amplitude of seismic wave recorded by local stations. For shallow earthquakes, surface waves are dominant in the seismic wave recorded by a station far away from epicenter and B. Gutenberg suggested determining the surface wave magnitude M_S from the maximum horizontal displacement of a surface wave at 20 second period. The body wave magnitude M_b is determined from the maximum amplitude of a body wave and its corresponding period. M_b is generally determined for the deep and long distance events.

One or more of these three magnitudes appear in most catalogs, such as those from the Geological Survey of Canada and the National Earthquake Information Service in the U.S. The relations among them are often given empirically as

$$M_S = 1.59\, M_b - 4.0$$

$$M_S = 1.27\, (M_L - 1) - 0.016\, M_L^2$$

For a more detailed discussion see Xu and Zhou (1982) and Aki and Richards (1980).

Since any magnitude is calculated from a maximum amplitude of a seismograph, the question of the physical meaning of magnitude arises. Intuitively the magnitude depends on the energy released through seismic waves, but it is difficult to relate theoretically to other important parameters. A magnitude is calculated from seismic wave amplitude at a given period while seismic energy release involves integration over the whole spectrum. This comparison and in particular the relation between the surface wave magnitude and the total energy E of seismic waves has been studied by many investigators. Gutenberg and Richter (1956) obtained an empirical relation

$$\log E = 1.5\, M_S + 11.8$$

Kanamori and Anderson (1975) suggested that this relation results from assumption of a unilateral propagating fault model together with a similarity assumption for source parameters. Duda (1978) suggested that the magnitude at period T should be a measure of the average energy flux density at this period. The average energy flux density $\overline{W}$ is related to M_b by

$$\overline{W} = D10^{2M_b - 6.0}$$

where $D=1/2\pi^2\rho\alpha$ is a constant , ρ and α are the density and the wave velocity respectively.

Seismicity can be broadly classified into foreshocks, mainshocks and aftershocks. Foreshocks and aftershocks appear to have a relationship to some particular mainshocks and occur before and after those shocks respectively. Foreshocks do not appear as clearly in catalogs as do aftershocks. The mechanism of foreshocks may be more complicated than that of aftershocks, but the same windows used for determining aftershocks also can be used for foreshocks when time is reversed and the magnitude of foreshocks is recognized as smaller than that of the mainshocks (Lamoreaux, 1982).

To a first approximation mainshocks are usually Poissonian distributed (Gardner and Knopoff, 1974) but recent studies on earthquake catalogs have shown that the non-independence of earthquakes is a world wide phenomenon at all magnitudes for earthquakes that appear on well-documented catalogs (Kagan and Knopoff, 1978, 1980,

1981a,b; Andrews, 1980, 1981). These studies show that a power law self-similar distribution of the time intervals of the nearest-neighbor pairs of earthquakes in a certain region. The spatial distribution of the distances between pairs of earthquake hypocenters, the areas of triangles formed by taking triples of hypocenters, and the volumes of tetrahedra of quadruples of hypocenters also show both a power law distribution and self-similarity.

There is no theoretical upper bound on the magnitude of earthquakes that appear in catalogs but practically it is limited because of the finite strength of rocks. A lower bound of the magnitude of recorded earthquakes comes from the detecting threshold of seismometers and the space coverage of the observing networks. Consequently the time and space patterns of earthquake occurrence can be distorted by changes of completeness of the seismicity catalogs at lower magnitudes. This means that studies of correlations between large and small events above some magnitude M must be restricted to those events that occur after a catalog becomes complete for the events larger than magnitude M. Unfortunately the idea of completeness of a catalog is not well defined. In practice it can be tested by considering whether the statistical behavior of small earthquakes is same as intermediate ones and whether the additional space coverage of observing networks will change the statistical behavior of the catalog.

3 Aftershocks and Omori's Law

Aftershocks are the most obvious feature of an earthquake catalog. Two major characteristics of aftershocks are their existence in all catalogs and their decrement with time both in numbers of events and in energy being released. Because aftershocks concentrate both in time and space around mainshocks they are usually assumed to be local effects of mainshocks and physically related to its fracturing process. In the analysis of the evolution of lithosphere dynamics aftershocks are often excluded from catalogs. Nevertheless changes in aftershock behavior may be a precursor of a large earthquake (Keilis-Borok et al, 1980a,b).

The relation between the number of aftershocks and time is known as the Omori's law (Lamoreaux, 1982)

$$n[T(e)] = \frac{C}{[T(e) - \beta]^{\alpha}}$$

where n[T(e)] is the average number of aftershocks during the e-th day after the mainshock and C, α and β are constants. The usual values for α and β are 1 and 0. The coefficient C generally depends on the magnitude of the mainshock and the magnitude range in which the aftershocks are counted. Lamoreaux (1982) has shown that the number of aftershocks will not depend on the magnitude of the mainshocks, M_m, if the aftershocks are counted in a magnitude interval from M_m-Δ to M_m, where Δ is a constant. This result suggests a self-similarity which imposes an additional constraint on the study of the physical nature of aftershocks.

4 Self-similarity and Magnitude - Frequency Relation

Stochastic self-similarity appears in many natural phenomena, such as Brownian motion and turbulent flow. A stochastic self-similar object is such that the "picture" of it shows no characteristic length scale. This means that images examined at two different magnifications will have the same statistical distribution. Mathematically, a set is said to be statistically self-similar if it is congruent in distribution to itself under a self-similar transformation (Mandelbrot, 1983), which transforms the point $x = (x_1,...,x_E)$ into the point $r(x) = (rx_1,...,rx_E)$, where r is a real number, in a Euclidean space R^E.

The self-similar sets are described as fractal by Mandelbrot. Many objects in nature have fractal geometry. An important parameter of fractal objects is their fractal dimension. It is the exponent D that might relate, for example, the apparent length P of the coastline of England to the length l of a measuring stick used

$$P = l^{1-D}$$

Note that the apparent length of the coastline increases exponentially when the measuring stick become smaller[1]. Recent measurements on the topography of natural faults and

[1]The west coast of Great Britain has a fractal dimension of D=1.25 (Mandelbrot, 1967).

fractures show that the surfaces are fractal or nearly fractal over a broad band (Scholz and Aviles, 1985). They found that D from 1 to 1.26 for wavelength varying from micrometers to meters.

In the dynamics of the lithosphere self-similarity appears in both time and space in earthquake behavior. Besides the self-similar distribution of earthquake epicenters, recent research on microfracturing in rocks indicates such feature exists at the scale of rock sample size in creep experiments. Hirata, et al., (1987), found that the fractal dimension of the spatial distribution pattern in their experiment ranged from 1.75 to 1.25 depending on the stage of creep. The self-similarity in time of earthquake occurrence is reflected in the magnitude and frequency relation (Gutenberg and Richter, 1954)

$$\log N = a - b\,M$$

where N is the number of earthquakes with magnitude greater than or equal to the magnitude M; a and b are constants which vary for each region. The value of b can range from 0.5 to 1.5, but is usually between 0.7 to 1.0 for tectonic regions (Mogi, 1985).

Much theoretical and experimental work has been done to connect the value of b with other physical quantities. Although observations and experiments indicate some connections between b, the stress state of the region, and the properties of the materials, none of the theoretical models was satisfactory (Geng, 1986). In laboratory experiments, it is found that microfractures radiate elastic waves in a manner similar to earthquakes during the deformation of rocks. Mogi (1962a, 1962b) found that the magnitude-frequency relation of microfracturing events is the same as that for earthquakes. He demonstrated that in many ways the statistical behavior of microfracturing activity observed in experiments is similar to that which has been observed for earthquakes, and related the value of b to the degree of heterogeneity of the model material. Scholz (1968) found, however, that the state of stress, rather than the heterogeneity of the material, plays the most important role in determining the value of b. The study of acoustic emission in rock creep experiments by Mogi shows that b decreases with time under a constant stress state. He concludes that b is mostly determined by the properties of the model material (See the review by Geng, 1986).

The value of b often decreases temporarily prior to a large earthquake (Suyehiro, et al., 1964; Mogi, 1985) and this change has attracted attention as a possible precursory phenomena of large earthquakes. Scholz (1968) claimed that the change in b results from an environmental stress change prior to a large earthquake, but Mogi (1981) argued that even under a constant load, the value of b still gradually decreases before a major fracture. No matter what the mechanism may be, it is generally accepted that the value of b tends to decrease prior to a large earthquake.

If we replace the magnitude M in the magnitude-frequency relation by the energy E, we have a power law relation between N and E

$$\log N = a + 7.8b - \frac{2}{3}b \cdot \log E$$

The N-E curve is a straight line on a log-log plot. Such a power law relation is a universal phenomenon, which appears in many physical systems and has been studied for over fifty years. A classical study of the power law distribution of broken coal was carried out by Bennet (1936). Other fragmented materials, such as granite, basalt, asteroids and interstellar grains are also found to satisfy a power law relation (Turcotte, 1986). In the case of the electric current in metals or semiconductors, it is found that the power spectrum of the current depends on the frequency with a power law $1/f^{\alpha}$ (α generally $0.8 < \alpha < 1.4$) (Yu and Liu, 1983; Weissman, 1988), this is known as 1/f-noise.

Noticing the property of scale invariance of 1/f-noise, Mandelbrot argued that it can legitimately be called "scaling noise" since it is noise whose character remains unchanged as it is contracted or expanded in time (Mandelbrot and Voss, 1983). The mathematical similarity between fractal and 1/f noise remains clear, but the physical connection is obscure. It seems that there is no "universal" generating mechanism. However its near ubiquity suggests that at least some feature of explanation should not depend on detailed models of a particular material.

In some cases it may be related to chaotic behavior, for instance, a large class of maps which generate intermittent signals display 1/f noise (Schuster, 1988). In other cases it may due to a wide distribution of characteristic energies, such as 1/f noise in metal, and it may be related to the non-Markovian diffusion transport (Yu and Liu, 1983).

We have to notice that this self-similar picture has to be band limited. The fractal description of time-distance-magnitude patterns of earthquake occurrence is a mathematical approximation. For natural faults Scholz (1988) has found that geometrically unmated fractal surfaces will develop a characteristic length when in contact under a normal load, because long-wavelength apertures close under a load whereas short-wavelength apertures may remain open. In the magnitude-frequency relation we notice that either the logN-logE curve has to bend to decrease the slope at low energy end or to bend to increase the slope at the high energy end to ensure a finite rate of energy release in earthquakes. Laboratory acoustic emission experiments have shown that the slope decreases at the low energy end of the curve (Scholz, 1968).

The existence of self-similarity implies that there is no particular scale involved in the evolution of earthquakes and that therefore a model which is capable of describing the dynamics of the lithosphere should also be scale free. Mandelbrot argues that a linear system would preserve a scaling, i.e. fractal geometry, but some non-linearities are usually required to create fractal structures (Mandelbrot, 1983). The fractal time and space patterns of earthquake occurrence suggests, therefore that such a phenomenon results from the non-linearity of the dynamics of the lithosphere.

5 Some Earthquake Related Physical Models

The occurrence of an earthquake can be considered as a sudden rupture of a fault, but the mechanics of those rupture zones is not well understood. Most physical models of the failure process of rock incorporate some kind of weakening behavior of the fault zones. The effects of liquids contained in many fault zones on the strength of rock, the so called Rhebinder effect, was recently suggested as important in earthquakes (Keilis-Borok, 1990). It states that the liquids penetrate cracks to reduce their surface tension. The cracks then grow with the drops of liquid propelling forward. Due to this mechanism alone the effective strength of solid rock may drop by a factor 10^5.

Besides the instability due to the Rhebinder effect, constitutive models also contain instability which can be interpreted as an earthquake (Rice, 1983). Slip weakening has

been observed in laboratory shear tests and modeled by the finite element method (Dieterich, 1978, 1979), and later had been worked into a state variable description by Ruina (1980, 1983). He assumed the friction stress τ depends only on the slip rate V, the normal stress σ and the state θ. The rate of change of state is assumed to depend only on the instantaneous state, the normal stress and the state,

$$\tau = F(\sigma, V, \theta)$$

$$\dot{\theta} = G(\sigma, V, \theta)$$

where θ is a collection of variables θ_i. One state variable is usually adequate to describe experimental results; sometimes two are needed.

Figure 1a shows a slip weakening friction relation. The shear stress τ required for slip first increases to a peak τ_p with slip δ, then decreases to a residual value τ_r after an amount of relative slip δ_0 (Rudnicki, 1988). This model is a generalization of the concept of static and kinetic friction; instead of an abrupt drop from the static to the kinetic value with the onset of slip, there is a transition reflected by the decrease of shear stress from peak to residual value. Such models are rate-independent, they lack any mechanism for regaining strength after reduction to the residual level and, consequently, they can not simulate the earthquake cycle on the same fault segment.

The rate-dependence of friction was demonstrated by Dieterich (1978, 1979, 1981) and other geophysicists (Ruina, 1983; Weeks and Tullis, 1985). Figure 1b shows the velocity weakening friction relation. The sliding velocity suddenly increase from V_1 to V_2, the shear stress exhibits an instantaneous increase from $\tau_{ss}(V_1)$ and then evolves to a new steady state value $\tau_{ss}(V_2)$. The negative $\partial\tau_{ss}(V)/\partial V$ will generate instability.

An instability due to an explosive increase of temperature in faults has been proposed by Griggs and Baker (1969), and may be a mechanism for deep focus earthquakes (Ogawa, 1987; Mahboobi, 1981). In this model, the fault zone is a homogeneous non-Newtonian viscoelastic material with a spatially variable initial temperature distribution. The shear deformation in the material concentrates in a thin layer

that initially has the highest temperature. The temperature increase due to the dissipation causes melting and thus induces slip along the shear zone. The strength of the shear zone will be regained as it is cooled due to thermal diffusion. The entire process is a quasi-static one in which inertia is not included explicitly. The instability generally takes the form that the slip rate becomes unbounded on a portion of the fault. The model fails, or tremendous difficulty is involved in the modeling at the point when the inertial effect is not negligible.

In these models, an instability can be interpreted as an earthquake. The instability arises because of the geometry of the fault or the nature of the constitutive relation. An indication of the progress in understanding is the current general agreement that some kind of weakening behavior is needed for instability. These models, however, either lack mechanism for regaining strength or fail beyond the unstable point. They are also simplistic in terms of the geometric complexities of an actual fault system.

6 Lithosphere and Its Dynamical Evolution

Most earthquakes occur within the lithosphere, which consists of top layers of the earth and is composed of brittle materials. The development of plate tectonics has provided a method for visualizing the structure of the lithosphere. Boundaries of major tectonic plates, such as San Andreas fault on the west coast of the US and the subduction zone on the west coast of Canada, define the major seismic zones around the world. Within the plate, however, fracturing occurs on all scales, from continental to microscale. The study of Allegre et al. (1982) has shown that the lithosphere presents a hierarchy of volumes, or blocks, which move relative to each other. The size of the volumes range from tectonic blocks as an upper limit to the microstructure such as quartz grains as a lower limit. These results have led us to visualize the lithosphere as a fully cracked body with fractal geometry of fractures (Figure 2, Keilis-Borok, 1990).

Tectonic plate motion is associated with mantle convection, which acts as a strain energy supply to the lithosphere. The energy that enters the system escapes by earthquakes and aseismic creep of faults. The following simple model will help to understand the relation between tectonic driving and earthquake occurrence. If we know the average

displacement, $\bar{u}$, associated with earthquakes and the relative plate velocity v, the average recurrence time depends on the driving velocity v by (Sykes and Quittmeyer 1981)

$$\bar{T} = \frac{\bar{u}}{\alpha v}$$

where α is the ratio of seismic slip to total slip. Real data, however, show a high deviation from this mean value.

The high deviation possibly results from the complexity of fault mechanics and interactions between block boundaries. Traditionally a fault is regarded as a passive interface which is locked by friction or cohesion. Earthquakes are identified as the fractures and unstable sliding, stick slip for instance, occur in these fault zones under tectonic stress or local stress concentration. The healing processes that follow the rupture of the faults invoke a micro process or a chemical process.

The studies in last a few years, however, suggested a new conception of the boundary zones of blocks as an active interface (Keilis-Borok 1990). The boundary zones are in active control of motion due to numerous processes, such as interaction with fluid, influencing friction and cohesion. Therefore, while the energy of the motion is stored within the whole lithosphere and well beneath, the release of the energy, the motion, is controlled from within the boundary zones.

Considering the earthquake generating mechanism from the dynamical state of pre-existing faults will result in what seems a paradox; are those pre-existing faults caused by previous earthquakes? The conventional consideration of a single fault fracturing as an earthquake source has its limitation in the understanding of the dynamical development in source volume. An earthquake and the fault zone should be considered together as a result of dynamical development of the whole domain of the crust, which could be weakened through the failure of anastomosed faults in a lesser scale (Newman and Knopoff, 1982,1983; Knopoff and Newman, 1983). Therefore, the fractal geometry of fault distribution within the lithosphere can be considered, in this sense, as the result of the dynamical evolution of the lithosphere.

A few theoretical models have explored the effect of the complex fault interaction on earthquake occurrence in last decade. Gabrielov's model (1986) of a complex earthquake

source associated with a large number of sub-faults has shown the significant role played by the interactions within fault system. The observed radiation patterns of seismic wave seem to support the idea that there is more than one fault related to the earthquake source (Keilis-Borok, 1990).

Using the hypothesis of scale free distribution of fractures, a number of geophysicists have attempted to use renormalization group approaches to model the hierarchical fault structure (Allegre, at el., 1982; Smalley and Turcotte, 1985). These studies have shown that the catastrophic failure is the result of asperity failure cascading away from the nucleus of failure. Foreshocks and acoustic emission observed in laboratory tests also imply that fractures at the macroscopic scale are a consequence of rupture at lesser scales.

Tremendous difficulties arise from the fact that fractures exist in all scales within the lithosphere, while all the existing theories assume continuous media. As a result the question of how to model the dynamics of the lithosphere including seismicity is still open. As the most successful models have a small number of degrees of freedom, recent developments in theoretical modeling of seismicity assume that the number of degree of freedom in such models is small. Work done so far is mostly concentrated on the behavior of a single fault plane as discussed in the previous section and it explores the consequences of some particular aspects of the earthquake process in isolation from the complexities of competing mechanisms occurring in the earth.

When more than one fault segments is involved, the strong interaction between them will possibly lead to a chaotic behavior (Gabrielov, et al., 1986). As Keilis-Borok (1990) pointed out that the dynamics of lithosphere is governed by the interaction of its blocks across and along the hierarchy. This interaction is realized through the wide variety of mutually dependent mechanisms. Each of them creates instability. None can be singled out as a major one so that the rest can be neglected. The model which describes lithosphere dynamics should be able to represent directly the gross integrated characteristics of the lithosphere.

7 General Ideas of Premonitory Seismic Patterns

If gross integrated characteristics of the lithosphere exist, they should be reflected in

detectable patterns in seismicity. Lacking a better understanding of the dynamics of the lithosphere one has to seek ad hoc approaches to detect these patterns. One of such approaches is to monitor the patterns of seismicity that precede strong earthquakes to meet the need of earthquake prediction. Such an approach implies that the occurrence of large earthquakes may be signalled by the seismic behavior of zones which are very far away, and that some large events seem to be preceded by anomalous seismic behaviors (Keilis-Borok, et al., 1980a,b).

The physics behind those patterns is not well understood so far, but several qualitative theories propose some explanations. These include inhomogeneities in physical properties of the materials along a fault zone and the processes of crack population growth, independent of the material in which these cracks are formed (Lamoreaux, 1982; Knopoff and Newman, 1983). However, there is no satisfactory explanation which will account for most of these patterns.

Lamoreaux (1982) summarized these patterns as a combination of quiescence and activation, the former referring to a spatial-temporal reduction (gap) and the latter to an increase of seismic activity in comparison to the normal activity. Many of these patterns have been identified with the seismicity rate, the magnitude and the seismic energy released. For example, **Bursts of Seismicity** refers to three related seismicity patterns: **Bursts of Aftershocks**, **Swarm** and **Pattern Σ**. These three patterns consist of the abnormal clustering of earthquake in time, energy and space before a major earthquake. **Bursts of Aftershocks** occurs when a medium magnitude mainshock is followed by an anomalously large number of aftershocks concentrated at the beginning of the aftershock sequence. **Swarm** can be defined as a sequence of earthquakes which are close to each other in space, time and magnitude, and occurring at the time when the seismicity of the region is not below average. **Pattern Σ** is defined as a sharp increase in the sum $\sum E_i$, where E_i is the energy of an earthquake occurring within a limited magnitude range and within a sliding window. If those patterns are diagnosed, a strong earthquake will probably be expected during a period of 5 or 6 years in future.

A pattern recognition algorithms for intermediate-term earthquake prediction

(Keilis-Borok, et al., 1988; Brown, et al., 1989) consider different types of premonitory seismicity patterns such as abnormal clustering, activation-quiescence, strong variation of seismic activity in time. The characteristics of earthquake sequence are defined on large areas and time-windows, and represented by several functions which by no means are independent. These functions are mathematical description of these patterns, and they are informative at distinguishing two kind of objects: the Time of Increased Probability (TIP) for occurrence of a strong earthquake and the rest of the time domain. The remaining problem of diagnosing TIPs is reduced to pattern recognition.

The premonitory patterns preceding a strong earthquake can be a more general concept which includes those changes in geoelectric effects or shear wave velocity, but here it only refers to those patterns which can be diagnosed from earthquake catalogs.

1.8 Tectonic Regime and Seismicity of the Vancouver Island Area

I chose the Vancouver Island area, which has been very active in the last few million years, to test the theory developed in this dissertation. The tectonic regime of the area is dominated by the relation of the Pacific plate, the America plate and the Juan de Fuca plate (Keen and Hyndman, 1979). The northern part of the Juan de Fuca plate has been named the Explorer plate since it is active and moves independently. The boundary between the Explorer-Pacific plates and the Juan de Fuca-Pacific plates is defined by numerous en-echelon spreading axes, offset by short transform segments. The full spreading rate of the motion across this boundary ranges from about 4 to 6 cm/yr. The boundary between the Juan de Fuca and Explorer plates is known as the Nootka fracture zone, which is a strike slip fault with a left lateral motion of about 2 cm/yr, extending northeasterly from the north end of the Juan de Fuca ridge to the continental shelf off north central Vancouver Island (figure 3).

The boundary between the America plate and the Juan de Fuca and Explorer plates is a zone of subduction whose underthrusting probably starts near the base of the continental slope. At present the Juan de Fuca plate moves relative to the America plate in the direction of $N35^0E$ with a convergence rate of approximately 3.5cm/yr. The Explorer

plate moves relative to the America plate in the direction of N6^{0}E. The convergence rate has been 1.4cm/yr during the last million years.

The seismicity of the west coast of Canada (figure 4) is of special interest to this study. The catalog of the Geological Survey of Canada and the National Earthquake Information Service in the U.S. contain some events with magnitude as small as 2, but not all small magnitude earthquakes are present. Milne et al (1978) pointed out that most of the seismicity on the western Canada margin (figure 4) has been concentrated along the three major boundaries, the Queen Charlotte-Fairweather fault system (Pacific-America plates), the offshore ridge-fracture zone system (Pacific-Juan de Fuca plates) and Vancouver Island-Puget Sound region (Juan de Fuca-America plates). It includes the lithospheric seismicity associated with the transition from the Pacific into the North America plate.

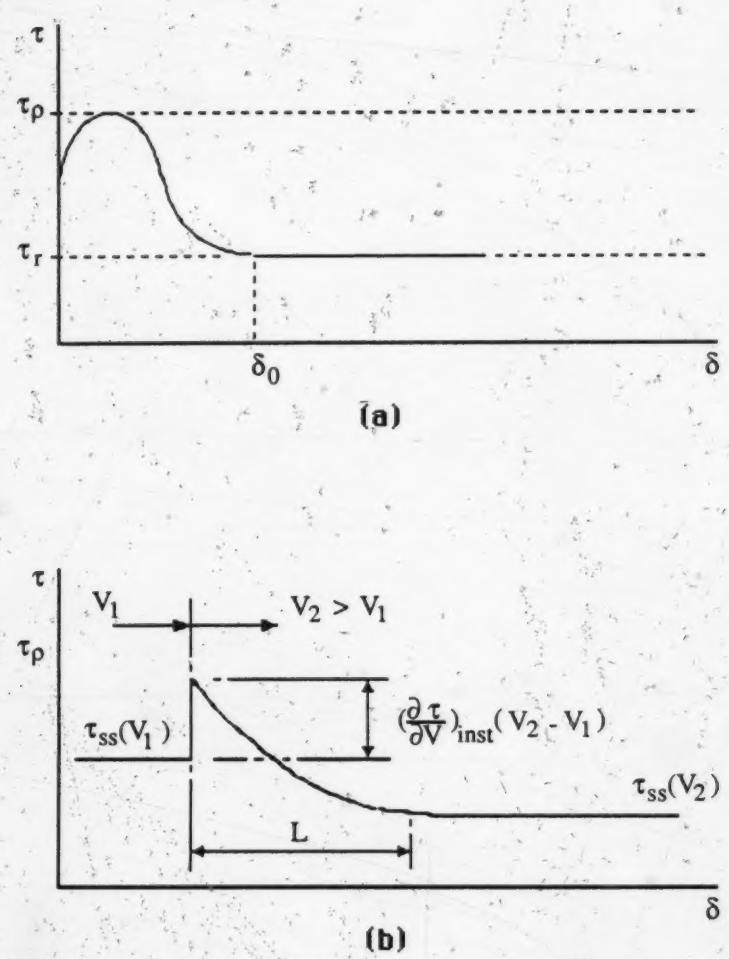


Figure 1 is an illustration of slip-weakening model (a) and velocity-weakening model (b).

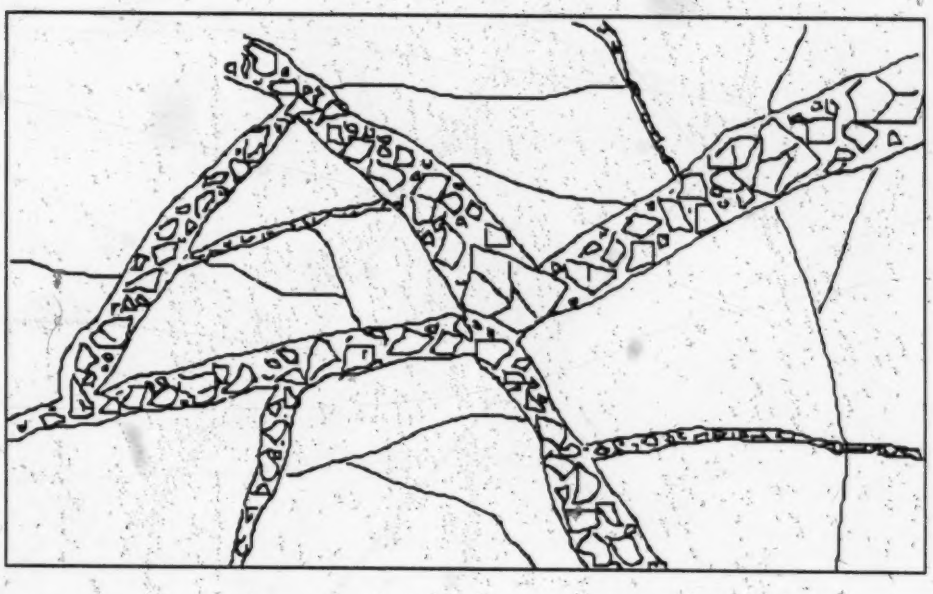


Figure 2 illustrates the idea that the lithosphere displays the hierarchical structure. Fracturing occurs on all scales, and has a fractal geometry of fault distribution (Adapted from Keilis-Borok 1990).

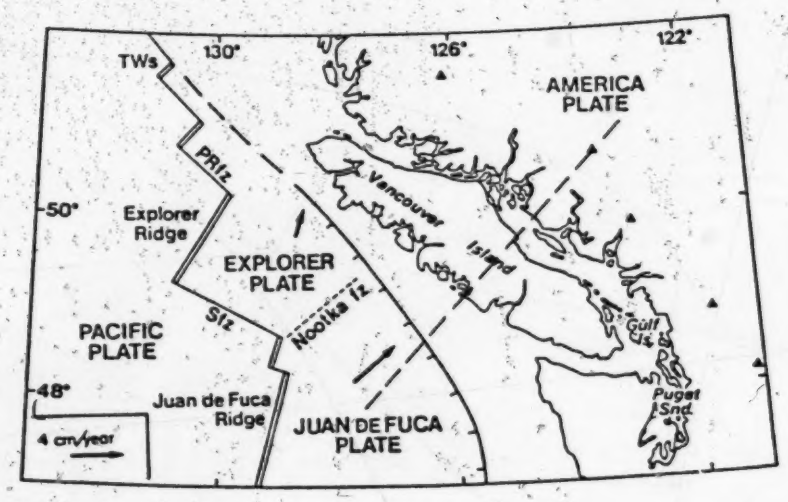


Figure 3 shows the main lithosphere plate boundaries and the relative plate motion of the west coast of Canada (Adapted from Li, 1986). The dashed line indicates the location of the cross section that governs our modelling in chapter 1.

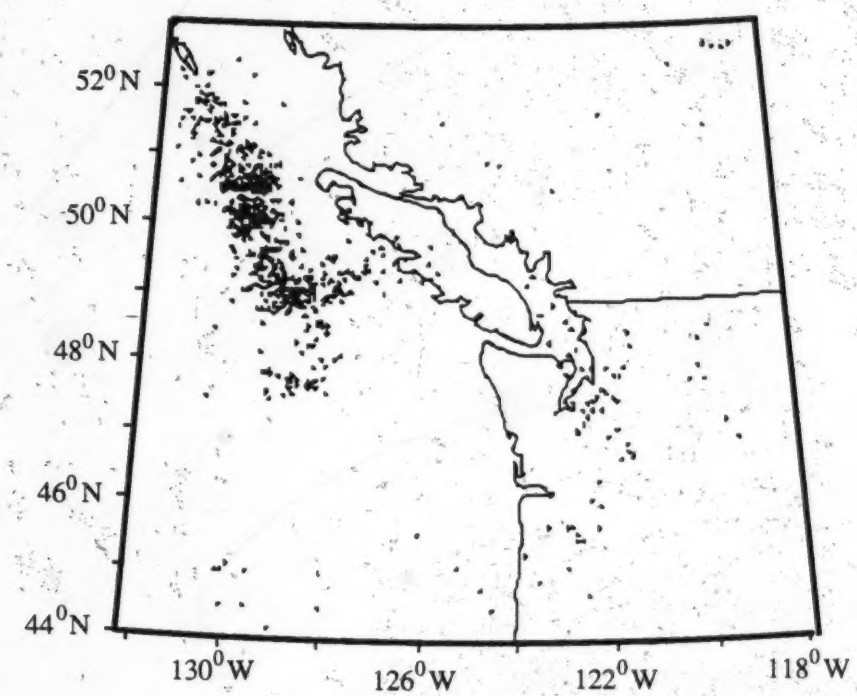


Figure 4 shows the geographic distribution of epicenters for events of magnitude 4 or greater on the west coast of Canada. Those events since May 1980 that were clearly related to the Mt. St. Helen's Volcanic Eruption were removed. The data mainly comes from the seismicity catalog of the Geological Survey of Canada and the National Earthquake Information Service in the U.S.

Chapter 1: A Shear Instability Model of Seismicity

Models are usually constructed in physics to complement phenomenology. Generally, they are constructed to represent certain behaviors of a physical system while not involving all its possible variables, and hence the complexities of the real system are greatly reduced. The simplicity is essential in modeling. A simple model may be so crude as to have little predictive power, however, it still manage to have its value in an explanative fashion, and serve as a starting point for more complete description.

One such model is an elastic finite element model of the Cascadia subduction zone including a subduction fault with nonlinear, temperature dependent rheology. It produces almost no conclusive results about the seismicity in the Cascadia subduction zone due to the uncertainties involved in the parameters of nonlinear rheology and the over simplified subduction zone geometry, but the way the model itself depends on a few control parameters is interesting. It suggests a relation between rheology of a fault zone and shear instability and earthquake generating mechanism in general, rather than the seismicity in the specific region.

1.1 Thermally activated instability: a quasistatic model

Griggs and Baker (1969) first investigated thermal instability of a creep process with a view to its applications to deep focus earthquakes. They demonstrated that if a material undergoing deformation has a creep rate that increases exponentially with temperature, the temperature rise would tend to accentuate any initial spatial inhomogeneity in the rate of deformation. Under certain boundary conditions of deformation this accentuation of strain rate may lead to a very rapid temperature rise and possibly melting. By introducing this idea to explain an occurrence of an earthquake in a subduction zone, we have the following picture: as a subducting slab, which is homogeneous and non-Newtonian viscoelastic, continues to plunge beneath another plate, the ductile part of the deformation induces heating due to viscous dissipation. If thermal conductivity is low enough, an explosive increase in temperature occurs due to a shear instability. The

temperature increase causes melting and thus induces slip along the shear zone, which is a thin region in which the deformation concentrated. The occurrence of the slip defines an earthquake. After the earthquake, the shear zone solidifies due to cooling caused by thermal diffusion and prepares for next earthquake. Figure 1.1 illustrates the idea that the shear instability is considered a mechanism of earthquakes. Ogawa (1987) has demonstrated the existence of such conditions for the occurrence of instability in subducting slabs. He demonstrated that the instability will occur for a wet dunite when stress exceeds 140 MPa based on the creep law of wet dunites determined by laboratory experiments at a temperature 1000^0K.

The melting, however, does not necessarily lead to an earthquake for melting on a fault surface is not always observed either in in-situ observation or in laboratory experiments. Nevertheless, the exponential increase of the temperature is still a possible way of triggering an instability, and this may be associated with some weakening mechanism at the microscale such as grinding of the materials in a shear zone (Mahboobi, 1981). The words *thermally activated instability* include the instability which might be caused in microscale and activated by the increase of temperature without melting.

This process was studied with a one dimensional model shown in Figure 1.2a (Mahboobi, 1981; Ogawa, 1987). The model is a layer of a viscoelastic material of thickness $2a$ embedded between two elastic outer layers with thickness $2(L- a)$. The viscoelastic region is more deformable, and can be interpreted as fault gouge or shear zone. The instability will take place within this region. The elastic region is taken to be of the order of the size of fault planes; the elastic energy stored in this region will be released as an energy source for shear heating when the instability takes place in shear zone. Such a model can be represented more clearly with a spring and a dash pot model and its equivalent form in which the effective elastic constant is used (Figure 1.2b).

The stress versus strain rate relation of the model is assumed to be

$$\dot{\varepsilon} = \frac{1}{Y}\frac{d\tau}{dt} + \dot{\varepsilon}_v \qquad (1.1)$$

where Y is the elastic modulus, τ is the shear stress, ε is the total strain and ε_v is the

viscous strain in the shear zone. For a long duration of application of stress, the creep rate in most rock depends on temperature and pressure by (Weertman, 1970)

$$\dot{\varepsilon} = D_0 f(\tau) e^{\frac{AT_m}{T}} \tag{1.2}$$

where D_0 and A are material parameters, T_m is the melting temperature, and $f(\tau)$ is a function of the shearing stress or deviatoric stress τ. Since T_m varies with pressure, this relation is also pressure dependent. $f(\tau)$ behaves linearly at a low shearing stress τ. A specific atomic model that results in a linear function of $f(\tau)$ could be the diffusive process known as Nabarro-Herring or Coble Creep. At a higher stress where solidus creep is more readily measurable $f(\tau)$ is a non-linear function of τ generally taken as a power law of the form τ^n, with n in the range of 2 to 5. When $\tau > 10^{-3}\mu$, where μ is the modulus of quasi rigidity governing rapid deformation of material, the power law breaks down and a more rapidly varying function of τ has to be considered (Weertman and Weertman, 1975). At such a high shearing stress, an exponential dependence of $f(\tau)$ has been suggested (Dorn, 1955; Van Bueren, 1961; and others). An empirical constitutive equation that represents these different regimes of stress dependence below T_m and at the same time has some physical basis may be written as (Mahboobi, 1981)

$$\dot{\varepsilon} = c_0^{-1} e^{\frac{-E}{kT}} \sinh\frac{v\tau}{kT} \tag{1.3}$$

where c_0 is a parameter related to vibration frequency of crystal lattice, E is the activation energy, v is the activation volume, k is the Boltzman's constant, and T is the absolute temperature. The hyperbolic function $\sinh\frac{v\tau}{kT}$ is proportional to τ at low shearing stress and exponentially dependent on τ at high shearing stress. It also has the right symmetry with respect to changes of sign of τ. A general physical basis for the form of equation (1.3) can be given in terms of thermally activated dislocation movement that becomes biased in a particular direction by an applied shear stress. In this case the processes of dislocation intersection which restrict the motion of dislocations are characterized by a stress-dependent energy of activation $U(\tau) = E - \tau v$ (Van Bueren, 1961; Feltham, 1966),

which obviously must remain positive.

Under a constant stress τ, the temperature profile in this system, which evolves in a quasi-static way, can be found as a solution of the conductive heat transfer equation with distributed heat sources per unit volume.

$$\rho c_p \frac{\partial T}{\partial t} - K\frac{\partial^2 T}{\partial x^2} = \tau\dot{\varepsilon} = \text{distribution of heat} \qquad (1.4)$$

where ρ is the density, c_p is the specific heat, K is the conductivity, x is the direction of heat flow, and t is time. The right side of equation (1.4) represents the heating due to viscous dissipation. The other heat sources, such as radiogenic heat source, are not included in the formulation.

Assuming a constant drive U at the boundary of the outer layers, we obtain the following quasi-static equation,

$$\int_0^L \dot{\varepsilon}\, dz = U\ .$$

Together with the boundary conditions for temperature,

$$T = T_0\ , \qquad \text{at } z = L$$

$$\frac{\partial T}{\partial z} = 0\ , \qquad \text{at } z = 0$$

equations (1.1), (1.3) and (1.4) form an initial value problem that can be solved in a routine way.

This development assumes both that inertial effects are negligible and that stress is constant over the thickness $2L$ of the fault. If either condition is violated we can say that an instability has occurred and this can be interpreted as a seismic event. Let μ be the shear elastic modulus of the fault material. If the creep properties of the medium change significantly in times of the order of shear wave travel time across the fault, $\frac{2L}{\sqrt{\mu/\rho}}$, inertial effects must be included. The criterion given by Mahboobi (1981) to indicate the moment at which the quasi- static model breaks down is

$$\rho \dot{v} \geq \frac{\tau}{L}$$

where v is the spatially integrated strain rate across the shear zone $\int_0^a \dot{\varepsilon}\, dz$. In this case inertial reactions are beginning to dominate applied forces and waves must be generated; we perceive a rupture to be occurring.

1.2 A single degree of freedom autonomous system

The model discussed in the last section is described by a set of partial differential equations (PDE's). For the purpose of simplicity in mathematics, it is useful to reduce the PDE's to a set of ordinary differential equations (ODE's) under certain approximations while the basic physical properties of the system still remain. Since temperature dependent viscous shear occurs in a very thin layer of thickness $2a$, compared with the tectonic scale, it is reasonable to approximate the fault zone by an infinitely thin fault. If the anelastic displacement at $\pm$ a is $\pm \Delta$, the shear strain rate $\dot{\varepsilon}=\frac{1}{2a}\frac{d\Delta}{dt}$. Let $\tilde{c}_0 = \frac{c_0}{2a}$; the time rate of the change of anelastic displacement can be written in terms of the stress and temperature by writing equation (1.3) as

$$\frac{d\Delta}{dt} = \frac{1}{\tilde{c}_0} e^{\frac{-E}{kT}} \sinh \frac{v\tau}{kT} .$$

When an instability takes place, the temperature will increase explosively within a very thin layer (Ogawa, 1987). Therefore, it is reasonable to consider the explosive temperature increase occurs within the shear zone, while the temperature distribution in the outer layers is constant. With this approximation, we can represent the partial derivative $\frac{\partial^2 T}{\partial x^2}$ by a finite difference scheme. Let T be the temperature at the center of the layer and T_0 be the ambient temperature, equation (1.4) becomes

$$\rho c_p \frac{dT}{dt} = \tau\dot{\varepsilon} - 2\frac{K}{a^2}(T-T_0) .$$

Defining σ as the mass per unit area of this fault, and $K_1 = 4K/a$ as the thermal conductivity for a thin fault we have

$$\sigma c_p \frac{dT}{dt} = \tau \frac{d\Delta}{dt} - K_1(T-T_0)$$

A similar simplification can be done in equation (1.1). If the total shear displacement at $\pm a$ is $\pm u_s$, the shear strain rate is $\dot{\varepsilon} = \frac{1}{2a}\frac{du_s}{dt}$. Defining $Y_1 = Y/2a$ as the elastic modulus for the thin fault, we have the stress versus strain relation of the shear zone:

$$\dot{u}_s = \frac{1}{Y_1}\frac{d\tau}{dt} + \Delta$$

When we introduce an effective elastic modulus $\tilde{Y}_1$ defined as $\frac{Y_1 Y_e/L}{Y_1 + Y_e/L}$ to replace Y_1, where Y_e is the elastic modulus of the elastic region, the displacement rate of the shear zone can be expressed in terms of the displacement rate at the boundary of the layer, U, which is assumed to be a constant. The loading rate of an external load on the fault is given as $\dot{f} = \tilde{Y}_1 U$.

We assume temperature fluctuation in the shear zone will not affect the ambient temperature distribution. As a result the temperature dependent creep behavior of a thin, uniform, fault element is fully defined by

$$\frac{dT}{dt} = \frac{\tau}{\sigma c_p \tilde{c}_0} e^{\frac{-E}{kT}} \sinh\frac{v\tau}{kT} - B(T-T_0). \qquad (1.5)$$

$$\frac{d\tau}{dt} = \dot{f} - \frac{\tilde{Y}_1}{\tilde{c}_0} e^{\frac{-E}{kT}} \sinh\frac{v\tau}{kT} \qquad (1.6)$$

where $B = K_1/\sigma c_p$.

A heterogeneous fault can be coarse grained into a number of fault segments. The interaction between faults elements is mediated by the surrounding medium. This effect can be treated by a finite element technique, and the stress equation (1.6) can be described in a matrix form (see section 1.4). The simplest case is the system with only one fault element. Its dynamics is described by equations (1.5) and (1.6). The system is autonomous since the right sides of the equations do not contain the time explicitly. Such a

system is often referred as a single degree of freedom autonomous system, and its general form is

$$\dot{x}_1=F_1(x_1,x_2)$$
$$\dot{x}_2=F_2(x_1,x_2)$$

Quite often, in the case of a system of nonlinear equations, no explicit solutions in the form of time dependent functions can be expected and we must be content with a statement concerning the stability characteristics of the system in the neighborhood of known motions. Of particular interest is the representation of a dynamical system in a phase space, which is a 2-dimensional space with x_1, x_2 as coordinates for a single degree of freedom autonomous system, and in general an N-dimensional space for a system described by N differential equations. Each point in phase space corresponds to a definite state of the system, and the motion of this point describes the time evolution of the system. Geometric analysis of the integral curves provides a useful tool in the study of phenomena associated with autonomous dynamical system. In particular, it provides considerable insight into the nature of the motion when our interest lies in the qualitative aspects of the motion rather than in solutions in the form of explicit functions of time.

For a single degree of freedom autonomous system the trajectories of the equations in the phase plane can be obtained by integrating the equations, with exception of the singular points where the trajectories passing through an ordinary point cannot approach it in finite time. The singular points are the fixed points for the flow, and the integral curve passing through a singular point consists only of the point itself. The fixed point condition is given as

$$\mathbf{F}(\mathbf{x}) = 0$$

where $\mathbf{F}=(F_1,F_2)$ and $\mathbf{x}=(x_1, x_2)$. Examining the property of the system at the neighborhood of a fixed point will help us to understand the dynamics of the system.

The geometry of the phase portrait of a single degree of freedom autonomous system is characterized by either an attracting point or a limit cycle. A stable steady state of

a physical system is represented by an attracting fixed point of the corresponding dynamic system, while an unstable state is represented by a repelling fixed point. In the neighborhood of the fixed point, the system can be linearized. The system near the fixed point is then described as

$$\dot{\mathbf{x}} = [\mathbf{J}]\,\mathbf{x}$$

where [J] is a Jacobian. The behavior of the system near the fixed point depends on the eigenvalues of [J]. The equilibrium state of the system is asymptotically stable if the real parts of all eigenvalues are negative. If p and q are the trace and the determinant of the Jacobian [J] respectively, we have eigenvalues:

$$\left.\begin{matrix}\lambda_1 \\ \lambda_2\end{matrix}\right\} = \frac{p}{2} \pm \sqrt{\left(\frac{p}{2}\right)^2 - q}$$

Different regions in the p-q plot will correspond to different behavior of the system (Meirovitch, 1970; Thompson and Stewart, 1986). Figure 1.3 shows such a classification, it characterizes the motion, stable or unstable. In the entire left half-plane the motion is unstable, it is characterized the saddle-point (SP) type of instability. In the right upper half-plane the motion is of the unstable node (UN) and unstable focus (UF) types, and in the right lower half-plane the motion is of the stable focus (SF) and stable node (SN) types.

If the eigenvalues cross the imaginary axis, the system will lose its stability. Suppose the dynamical system depends differentially on a real parameter μ (the bifurcation parameter). If the Jacobian is invertible, the implicit function theorem implies that the system should also have a unique fixed point when μ moves away a little from its original value. However, as μ is varied, the invertibility condition may eventually break down in several ways leading to points of changing stability.

When two eigenvalues cross the stability boundary to the right half complex plane as a complex conjugate pair, the system undergoes Hopf bifurcation. It replaces a steady state by a periodic state, characterized by a small closed curve. Figure 1.4 shows the phase portrait of Hopf bifurcation in the Guckenheimer-Holmes model.

There exist two parameters in the thermally activated instability model: the loading

rate which could correspond to a tectonic drive, and the ambient temperature which represents the thermal regime of the fault zone. In the numerical investigation of how the behaviors of the model depend on the control parameters, it is natural to reduce equations (1.5) and (1.6) to their non-dimensional forms by normalizing t, τ, and T by 1/B, $\tilde{Y}_1 L$, and $\frac{\tilde{Y}_1 L^2}{\sigma c_p}$. Here we introduce a length scale L which is related to the typical length scale of a fault system. After this normalization, equation (1.5) and (1.6) become

$$\frac{dT}{dt} = C\tau e^{\frac{-Q}{T}} \sinh\frac{A\tau}{T} - (T-T_0) \tag{1.7}$$

$$\frac{d\tau}{dt} = \dot{f} - C e^{\frac{-Q}{T}} \sinh\frac{A\tau}{T} \tag{1.8}$$

where t, τ and T are non-dimensional variables and will be used as such in the following sections. $\dot{f}$ is non-dimensional loading rate. C, A, and Q are non-dimensional constants defined as: $Q = \frac{\sigma c_p E}{k\tilde{Y}_1 L^2}$, $A = \frac{\sigma c_p v}{kL}$, and $C = \frac{1}{BL\tilde{c}_0}$.

In the above approximation, the quantities K_1, σ, and $\tilde{Y}_1$ are defined with respect to the effective fault thickness $2a$. In order to obtain the finite shear displacement across the infinitely thin fault, the elastic constant $\tilde{Y}_1$ for the thin fault must remain finite, i.e. the limit $\lim_{a\to 0} \frac{Y}{a}$ remains finite. Y_1 is so determined that displacement across the infinitely thin fault is of the same order of magnitude as the displacement across a length of tectonic scale. Since the stress in the infinitely thin fault and in its neighborhood are of the same order of magnitude, $Y_1 = Y_e/L$, where Y_e is the elastic constant of the neighborhood of the thin fault. In this case the effective elastic constant $\tilde{Y}_1 = Y_1/2$. We choose L=35km in our modelling; it is the typical width of fault zones and is also a coarse grain characteristic size of the fault geometry and layer size.

The material properties in the fault zone are largely unknown. The unquestioned existence of grains, fluids, and non linear processes leads to great uncertainty. The constants C, Q and A in this modelling are estimated from the typical physical constants in subducting slabs (Ogawa 1987). The density ρ is about 3000 kg/m^3, c_p about 800 J/kg deg, the thermal conductivity K is about 2 to 3 J/m-s-deg, the elastic constant Y is about 70 to 130 GPa, the activation energy E is about 390 to 570 kJ/mole. In our problem we deal

with processes on a much larger scale than atomic and estimate the activation volume as the square of the Burger vector of the dislocation times the jog spacing of dislocations (Feltham, 1966). This gives activation volumes on the order of 10^{-5} to 10^{-6} m^3/mole. The atomic vibration frequency $1/c_0$ is of order 10^{12} to 10^{14}Hz. The thickness of the shear zone $2a$ range from 0.1 to 1 km. In our one dimensional model, we choose the surface density σ by $\sigma/2a=\rho$. B is of the order of 10^{-11}, Q is taken as 0.05, A as 0.2. C is order of 10^{22} to 10^{23}.

The dependence of the stability of the system on the loading rate and ambient temperature can be investigated numerically. Figure 1.5 shows the relation between the real part of eigenvalues of the Jacobian and the loading rate. When the loading rate is small or large, the phase curves will be attracted to a fixed point in phase space. The system is characterized by a stable focus. With an intermediate loading rate, the phase curve is attracted to an invariant limit cycle. Hopf bifurcation occurs when the loading rate varies both from low to high around 0.001 and from high to low around 0.018, where the ambient temperature is taken as 7.8×10^{-4}. The existence of a stable region corresponding to high and low loading rate agrees with the results obtained from finite thickness fault model (Mahboobi, 1981). Figure 1.6 shows the dependence of the real part of eigenvalues of the Jacobian on the ambient temperature when the loading rate is taken as 0.01. The system is characterized by a limit cycle when the ambient temperature is low, and a stable focus when the temperature is high.

Figure 1.7 shows a limit cycle solution of the system calculated when the loading rate equals 0.01 and the ambient temperature equals 7.8×10^{-4}. After sufficient time, the system will settle down to such a periodic solution asymptotically. The figure shows that when the stress accumulates to a certain level, the temperature starts to increase exponentially due to viscous flow and the system become extremely unstable. At this moment the stress starts to decrease and the temperature increase. Such sudden drop in stress could be interpreted as an earthquake. Eventually the stress starts to increase due to tectonic drive, and the temperature starts to decrease due to thermal diffusion. It then leads to a new period of the motion.

When the system changes from focus to a limit cycle solution, it corresponds to an aseismic sliding of different fashion: stable and periodic. As the state of the system moves away from q axis in figure 1.3, the system become stiffer. The sudden drop in stress due to a temperature increase in this case will be so fast that the quasi-static approximation condition will be break down; it is associated with seismic wave generation. Figure 1.8 shows the stress change versus time for the model in which the loading rate is taken as 0.01 for both cases, and the ambient temperature is taken as $7.2\text{x}10^{-4}$ for the dash-curve and $7.0\text{x}10^{-4}$ for the solid curve. A simple criterion to indicate the condition where the quasi-static approximation breaks down is whether stress changing rate exceeds τ/T, where τ is 1kbar, the typical shear stress in subducting plate, and T is 0.1sec, the typical period of seismic waves.

1.3 The effects of the real fault gouge

In the model discussed in the previous sections, the fault gouge is approximated by a layer of homogeneous materials. The difficulty of the simulation arises when it is noticed that the stress at which the instability occurs is tens kilobars, while the generally accepted value at which a rock will fail is not over one kilobar. Such a big discrepancy can not be resolved simply by adjusting a few parameters in the system. Moreover, such a stress seems not to depend on most of the parameters in the system.

We know from the previous section that the fixed points play a very important role to the dynamic behavior of such a system, because they force a definite structure of the trajectories in their neighborhood. The system is stable in the neighborhood of an attracting fixed point in phase space, and unstable in the neighborhood of a repelling fixed point. In the latter case the instability will occur, which is interpreted as an earthquake. Therefore the stress at the neighborhood of a repelling fixed point can be used to approximate the stress at which the instability will occur. The stress corresponding to a fixed point is found as

$$\tau = \frac{T - T_0}{\dot{f}} .$$

It depends only on the loading rate and the difference between temperature at the shear zone and ambient temperature. The most remarkable feature of above expression is that it is independent of the creep law of the shear zone materials. This eliminates the possibility of reducing the stress by choosing a different creep law. On the other hand, the loading rate and temperature must be in a definite range in order to get a limit cycle solution and therefore, the choice for these parameters is very limited.

Since the need for a high failure stress can not be resolved by adjusting parameters in the model such a model is not realistic, and some important features of the fault gouge are ignored in this model. In order to lower the failure stress, the concept of effective stress is introduced as

$$\tau_e = \frac{\tau_a}{f}$$

where τ_a is an apparent stress, $0 < f < 1$, τ_e is an effective stress, which is the stress that actually participates in the mechanism of thermally activated instability. The factor f allows the effective stress to rise to one to two order higher than the apparent stress. One possible contribution to such a high stress within the fault gouge assumes that the fault is blocky, and we deal with the stress transmitted across a small contact area. In this case, the factor f equals the ratio of real contact area to apparent area. It is also well known that the presence of fluid in a fault gouge and the interaction between microscale fractures also play an important role in determining the strength of rock. How these effects contribute to the factor f is still not clear.

The value of f should be an increasing function of normal stress on the surface layer, however, this relationship is not very well supported by experimental observations. Mahboobi (1981) has introduced this factor in a shear instability model to account for the effect of presence of normal pressure (Figure 1.9). While the normal stress is below some threshold, the real area of contact depends on surface roughness, and the lateral interaction or interlocking of pre-existing asperities has to be taken into consideration. In the limit of the high normal stresses, f increases from its minimum value zero and approaches unity, therefore the region of contact will depend on a creep behavior similar to that of the surrounding material. This is a more satisfactory result than an indefinitely increasing

limiting frictional force inferred from the simplest extrapolation of the law of friction.

The effect of the factor f on the stability of the system is shown in figure 1.10. The sudden increase of the real part of the eigenvalue of the Jacobian suggests a sudden transition from stable state to unstable state as the factor f increases.

The jerky type of motion known as stick-slip is observed in many type of rock experiments and is often cited as a possible mechanism for earthquakes. The effect of confining pressure on the mode of sliding can be a promotion of a transition either from a stable sliding to a stick-slip or from a stick-slip to a stable sliding. Byerlee and Brace's (1968) experiments showed that stick-slip takes place when the confining pressure is high (~100 MPa). The results from Scholz, et al. (1972) and Ohnaka (1973) supported the effect of transition from stable sliding to stick-slip as confining pressure increases, but there is little agreement on the value of the pressure at which the transition takes place. However, Engelder (1974) showed the opposite effect of confining pressure on the transition. In his experiments on a sandstone rock, stick-slip was observed at confining pressure below 70 MPa and stable sliding above this value. The above seemingly contradictory observations are easily resolved with the result shown in the Figure 1.10; it predicts a stable sliding at both low and high confining pressure, and stick-slip in between.

1.4 The effect of inhomogeneity of a fault system

When a fault is inhomogeneous, different parts of the fault are generally in different states. One part of the fault will interact with other parts through the surrounding medium. In this case the fault can be coarse grained into a number of fault segments and the mean values over each fault segment are used to describe the state of the fault segments. This leads to a group of coupled ordinary differential equations. We know from the discussion of previous sections that the dynamics of each fault segment is governed by two ODE's, a stress equation and a temperature equation. It is reasonable to assume that the heat generated due to viscous flow in a fault segment mostly diffuses to its surrounding medium, and the heat transferred to its neighboring segments is negligible. The temperature equations therefore are not coupled to each other. When the stress state in one

fault segment changes, this effect will be transmitted to other fault segments by the surrounding medium. This implies that the coupling among the stress equations depends on the properties of the surrounding medium and its geological structure. The finite element technique is obviously a good candidate for exploring this problem.

The finite element problem for a model with discontinuities can be formulated as the solution of

$$\mathbf{Kd}=\mathbf{f}=\mathbf{A}_1\mathbf{b}+\mathbf{A}_2\delta+\mathbf{A}_3\mathbf{g} \tag{1.9}$$

where $\mathbf{K}$ is a stiffness matrix, $\mathbf{d}$ is the vector of N nodal displacements, and the source terms are contributions from boundary displacements $\mathbf{b}$, inelastic fault displacements δ, and gravity $\mathbf{g}$ respectively. The matrices $\mathbf{A_1}$, $\mathbf{A_2}$, and $\mathbf{A_3}$ are the transformations that convert their respective vectors into source terms or body force equivalents. $\mathbf{A_2}$ contains the effective shear modulus of the fault. In the actual process of computation, the terms on the right hand side arise from boundary constraints.

The slips are computed on fault elements by methods external to the elastic finite element calculation. The shear stress on a fault element is $\tau = \mathbf{M_1 d} + \mathbf{M_2}\Delta$, where $\mathbf{M_1}$ connects the displacements on the N nodes to the stress on the fault elements and $\mathbf{M_2}$ connects the inelastic displacements to the elastic stress on the fault elements. After replacing $\mathbf{d}$ with the solution of equation (1.9) and differentiating this with respect to time

$$\frac{d\tau}{dt}=\mathbf{M}_1\mathbf{K}^{-1}\left[\mathbf{A}_1\frac{d\mathbf{b}}{dt}+\mathbf{A}_2\frac{d\Delta}{dt}\right]+\mathbf{M}_2\frac{d\Delta}{dt}$$

The boundary conditions are

$$\mathbf{b}(t)=\mathbf{b}_0+t\sum_{k=1}^{K}v_k\mathbf{c}_k=\mathbf{b}_0+t\mathbf{C}\mathbf{v}$$

where the vectors $\mathbf{c}_k$ are composed of the indices of the nodes of all those boundary points that move at velocity v_k. In this case there are only K independent boundary conditions. The time rate of change of stress on the fault elements can be written

$$\frac{d\tau}{dt} = [\, M_1 K^{-1} A_1 C \,] v + [\, M_1 K^{-1} A_2 + M_2] \frac{d\Delta}{dt} \qquad (1.10)$$
$$= G_0 + G\frac{d\Delta}{dt}$$

The first term on the right side of the equation G_0 is the loading rate acting on the fault segments, and the coefficient of the second term, G, reflects the interaction among the fault segments. All the geological complexity of the fault region is reflected in these two matrices. These equations together with the temperature equations constitute a group of ordinary differential equations in the shear stress and temperature on the fault elements that completely determine the time evolution of the fault system. The evaluation of the matrices K, M_1, M_2, A_1, and A_2 is explained in Appendix A.

When the interactions among different fault segments enter the system, the dynamic behavior of the system becomes far more complicated than that of a single fault element. A simple case in this model involves two fault elements, where $G_0=(g_{01}, g_{02})^T$, $G = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}$, and the ambient temperatures are T_1 and T_2. Our first example uses $g_{11}=g_{22}=-1$ and $g_{12}=g_{21}=0.1$, the ambient temperatures are taken $T_1=T_2=7.8\times10^{-4}$. The loading rates g_{01}, and g_{02} will impose two basic frequencies on the system. As the loading rates change, frequencies locking occurs. Figure 1.11 shows the stress versus time for one of the fault segments when the loading rates are $g_{01}=0.01$ and $g_{02}=0.012$. The time sequence displays an erratic behavior. A further investigation of the one-dimensional map constructed by choosing successive maximum stresses indicates the motion is quasi periodic however. Figure 1.12 shows the one-dimensional map; it is a cross section of a two dimensional torus. At other loading rates the frequencies are locked to various rational numbers.

When we increase the interactions between two fault segments by increasing the parameters g_{12} and g_{21} the motion of the system become periodic. The system becomes very stiff when we decrease the ambient temperature, and this causes great difficulty in the numerical integration. It is uncertain whether the system will become chaotic at conditions of low ambient temperature and strong interaction between fault segments.

1.5 A model of the subduction zone near Vancouver Island, Canada

The two dimensional finite element model used in the example calculation discussed below is representative of the structure on a line crossing Vancouver Island and perpendicular to the coast of North America (Figure 3 in Introduction). Because the Juan de Fuca plate has moved in a direction N35° E relative to the America plate for the last few million years (Keen and Hyndman, 1979), we assumed the driving force on the Juan de Fuca plate would be in this direction. The geometric configuration of our model is mainly based on the results of analysis of LITHOPROBE seismic profiles (Yorath, et al., 1985; Green, et al., 1985, 1986) and the gravity model of that area (Riddihough, 1979). Since in a two dimensional elastic model, a bent fault must be accommodated by a more complex structure, we ignored the known sudden increase of the slope of the subducting oceanic plate beneath the continental plate under Vancouver Island and modelled the boundary between the subducting plate and continental plate and upper mantle which we considered a fault zone, by a straight fault of 18 degree dip.

Figure 1.13 shows the mesh of the model in which the material parameters are those used previously (Nyland and Li, 1986). The oceanic plate and the upper mantle beneath it move at a constant velocity along the fault. Such movement corresponds to the push from the Juan de Fuca Ridge or drag from mantle convection. The gravitational pull is not dominant in that area. The interface between the subducting plate and the continental plate is modeled by 10 non-linear visco-elastic elements. The surrounding materials are elastic and the fault elements interact with each other through this surrounding elastic media. The finite element model was used to calculate the interaction matrix and the loading rate on each fault element and the time evolution of the dynamics was explored for a variety of fault parameters.

The integration of the differential equations by numerical means posed some challenges. We used either a 5th order Gear backward differentiation method or a 12th order Adams predictor method as implemented in DIVPAG (IMSL 1987). At certain times the equations were clearly changing character, becoming stiff. Notice that the part of elastic energy stored in the medium will be radiated by seismic wave when an earthquake occurs, while in our model the inertial effect is neglected, and all the energy is converted to heat

which increases the temperature in the fault element. Therefore, it is reasonable to drop a certain amount of stress without increasing temperature to simulate the occurrence of an earthquake. This step is taken when inertial effects dominate. I believe that it is plausible, but cannot prove, that such a state corresponds to an earthquake.

Since the routine provides the option of controlling the minimum integration step size, we set this minimum to a value of the order of the time of transit for shear waves through the fault segment. When the routine sought to reduce the step size below this minimum we stopped the integration. We interpret this condition, which can be diagnosed as the approach of very rapid variation of the solution in time, as indicating that the inertial effect must be invoked and that failure is imminent. The equations discussed in the previous section will not be adequate to describe this failure. In this situation we determine which element has the highest rate of temperature increase. If the rate of change of stress is negative, we drop the stress on this element until the rate of change of stress become zero. Otherwise we raise the temperature of the element until the rate of change of stress become negative. The state after the stress drop is used as a new initial condition for the integration.

We chose the same fault parameters as those used in previous sections of this study for this model. The ambient temperature is taken about 700^0C for the top fault segment and 1200^0C for the segment at the bottom of the model. The ambient temperature for other fault segments are obtained through linear interpolation. The results of the calculation are shown in figure 1.14. The magnitude of each event is determined according to the stress drop associated with this event. The scale displayed in the figure 1.14 is arbitrary. The magnitude and frequency relation is shown in figure 1.15. Note that the curves have two quite distinct slopes. The stress drops resulting from successful numerical integration of the equations are more numerous than those that appeared when fracturing was invoked and may correspond to aseismic creep.

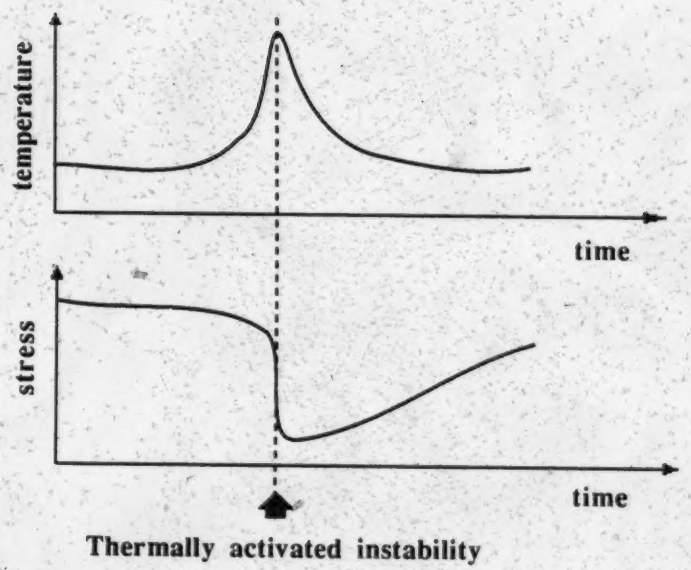


Figure 1.1 illustrates the idea that the thermally activated shear instability as a mechanism of earthquakes.

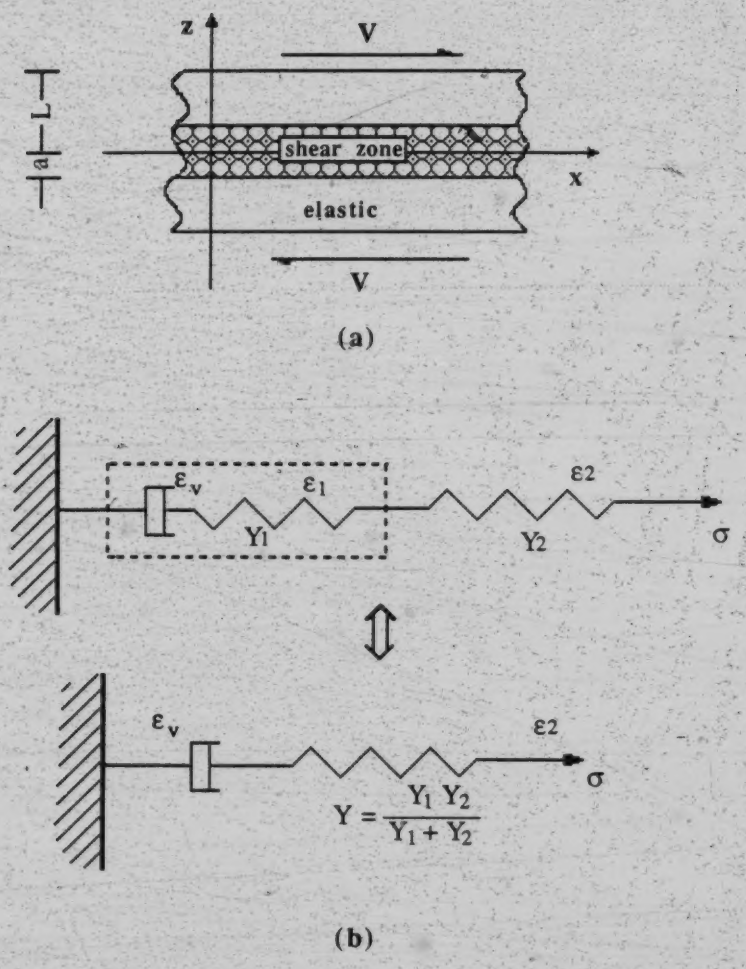


Figure 1.2 (a) illustrates a one dimensional model. (b) is its equivalent form described by a spring and dash pots.

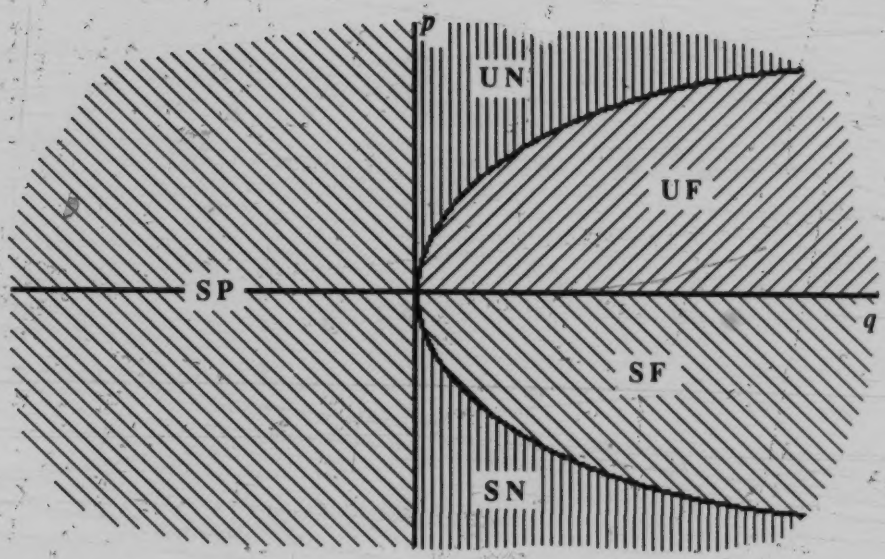


Figure 1.3 sketches the clarification of the phase portraits at the neighborhood of the fixed point.

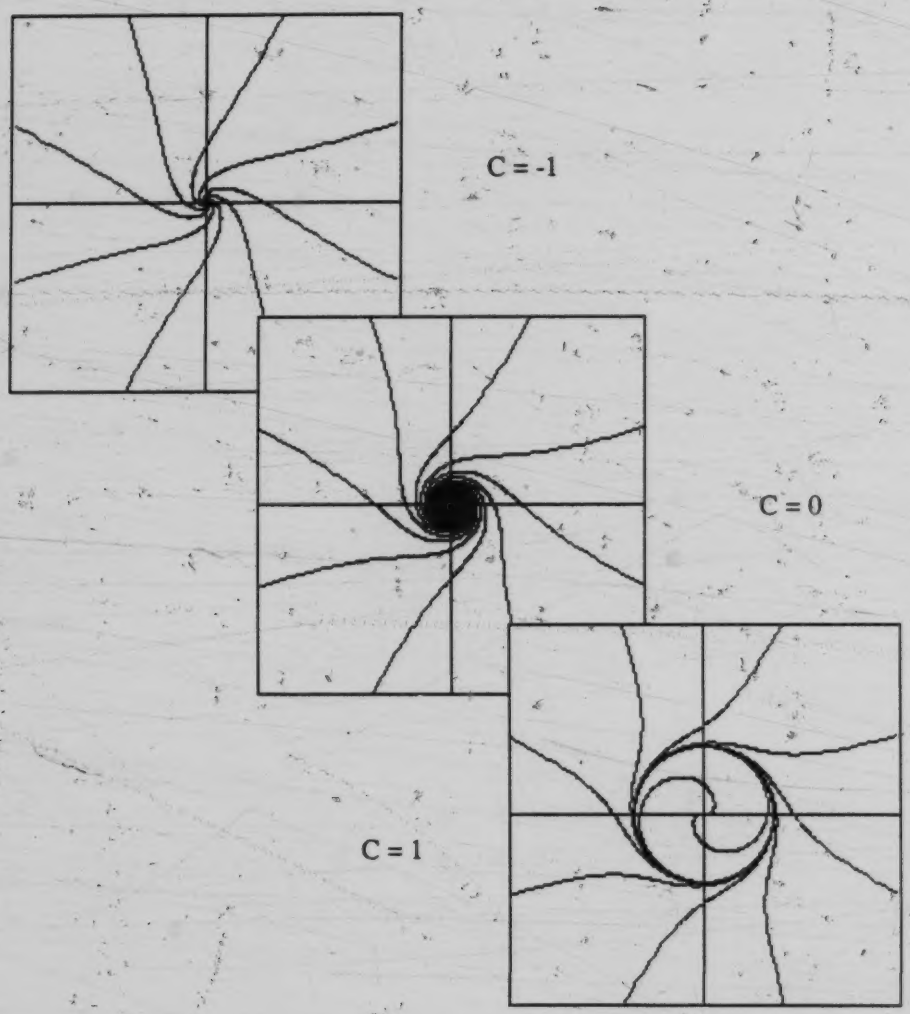


Figure 1.4 shows the phase trajectories of the system described by the differential equation $\dot{r} = r(c-r)$ and $\dot{\theta} = -1$. Hopf bifurcation occurs when the parameter C changes sign.

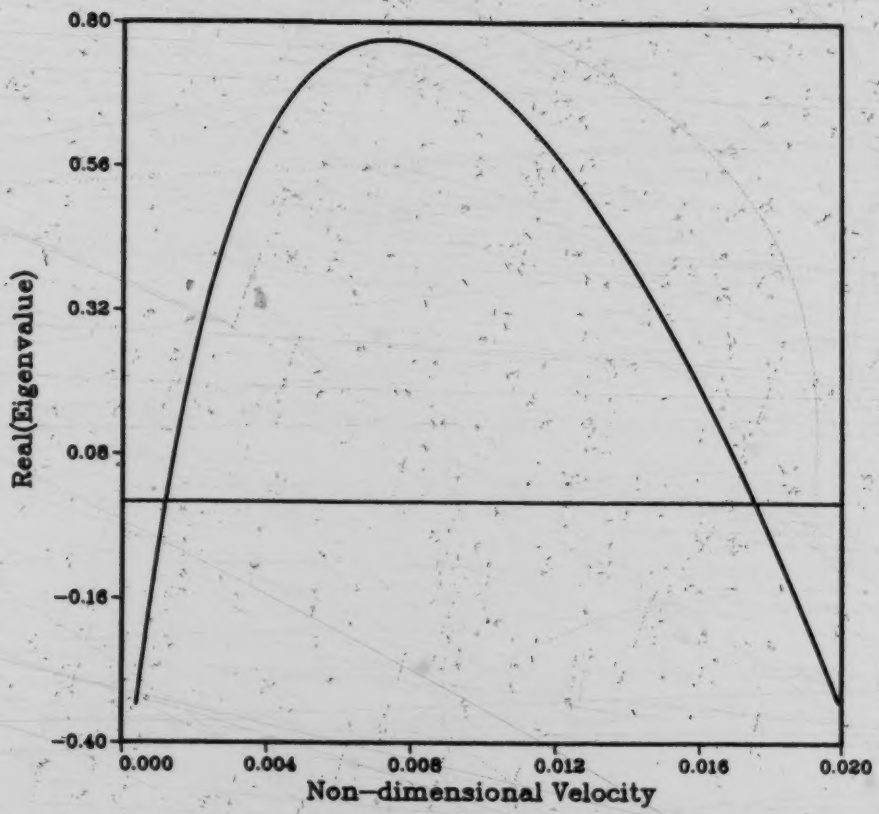


Figure 1.5 shows the relation between the real part of eigenvalue of the Jacobian and the loading rate. Hopf bifurcation occurs when the curve crosses zero.

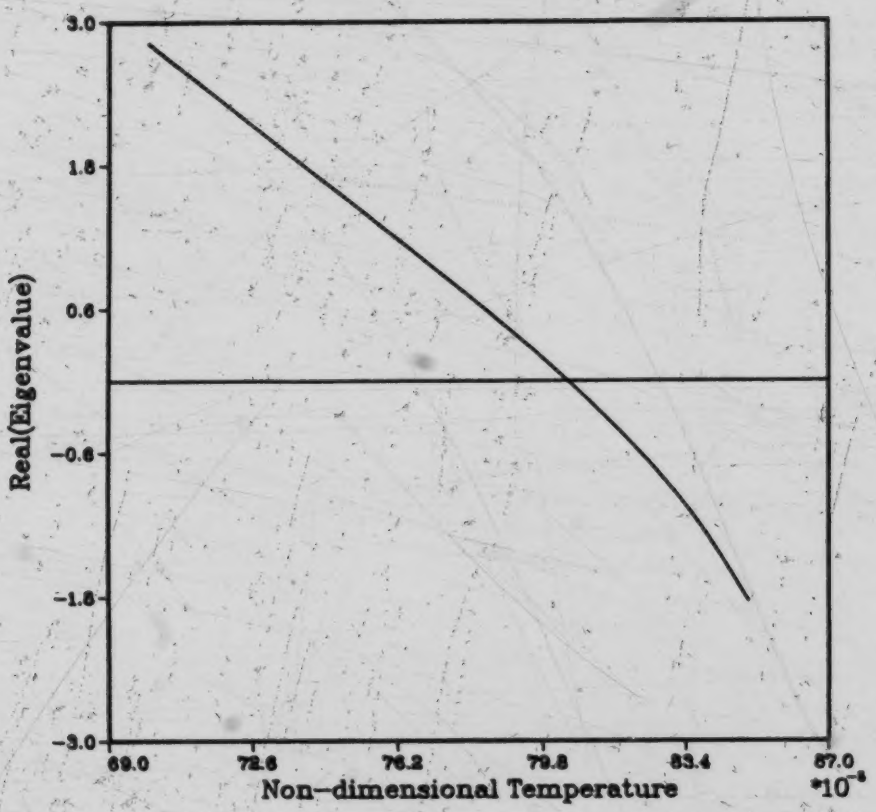


Figure 1.6 shows the relation between the real part of eigenvalue of the Jacobian and the ambient temperature.

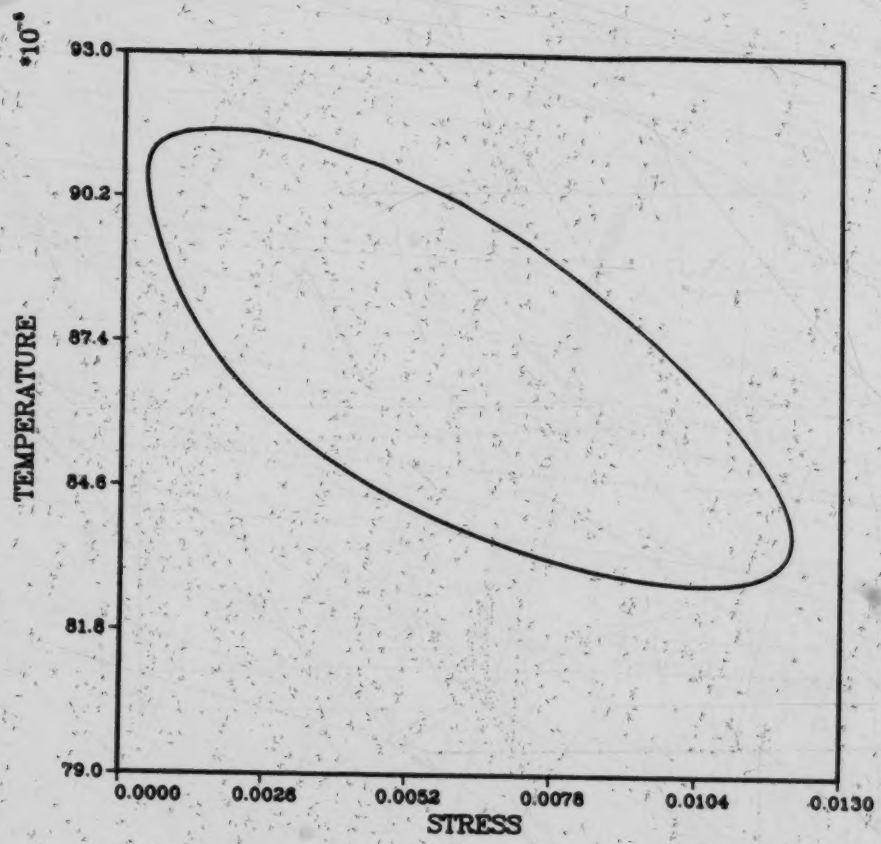


Figure 1.7 shows a limit cycle solution of the system.

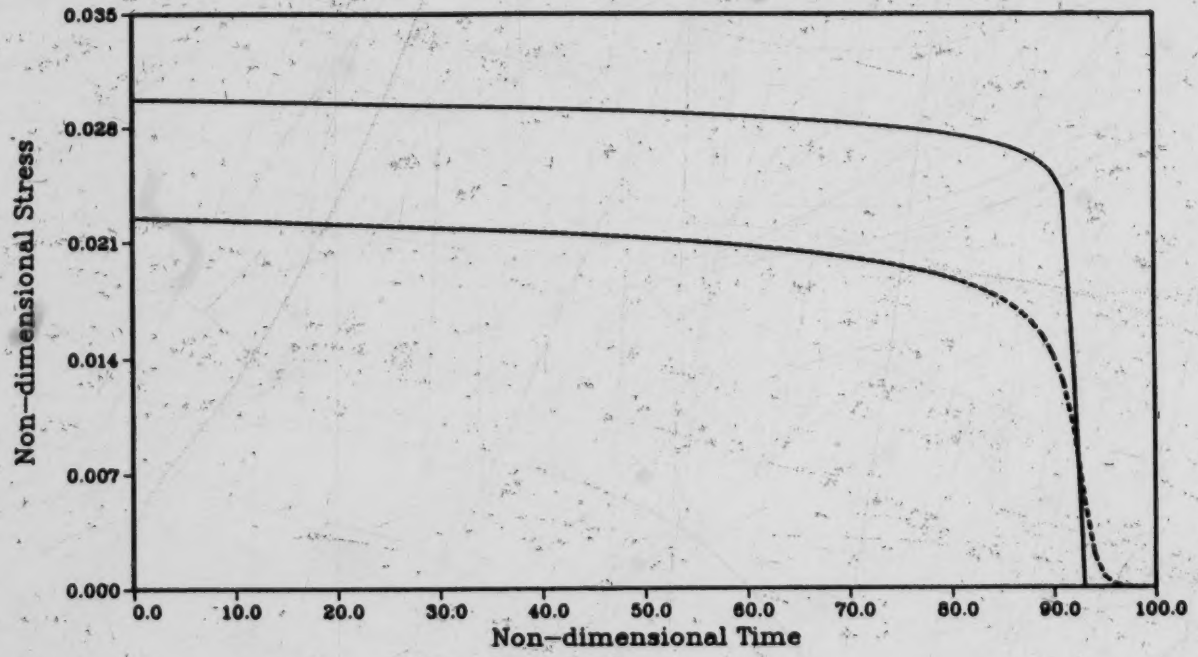


Figure 1.8 shows the stress change versus time for the model in which the loading rate is 0.01 and the ambient temperature is 7.2×10^{-4} for the dash-curve and 7.0×10^{-4} for the solid curve.

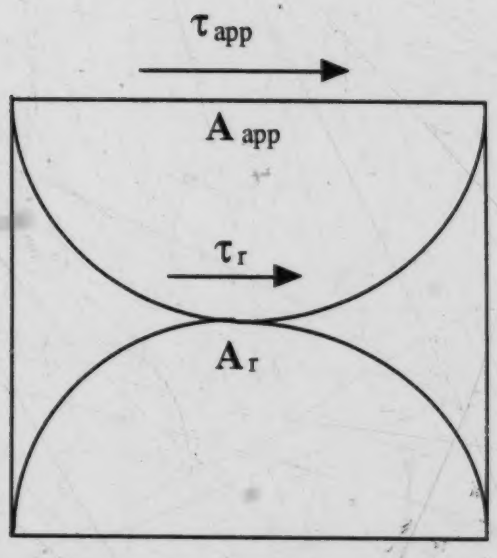


Figure 1.9 illustrates the idea that the real contact area is much smaller than the apparent contact area. The effect of normal pressure is taken into account by introducing the factor *f*.

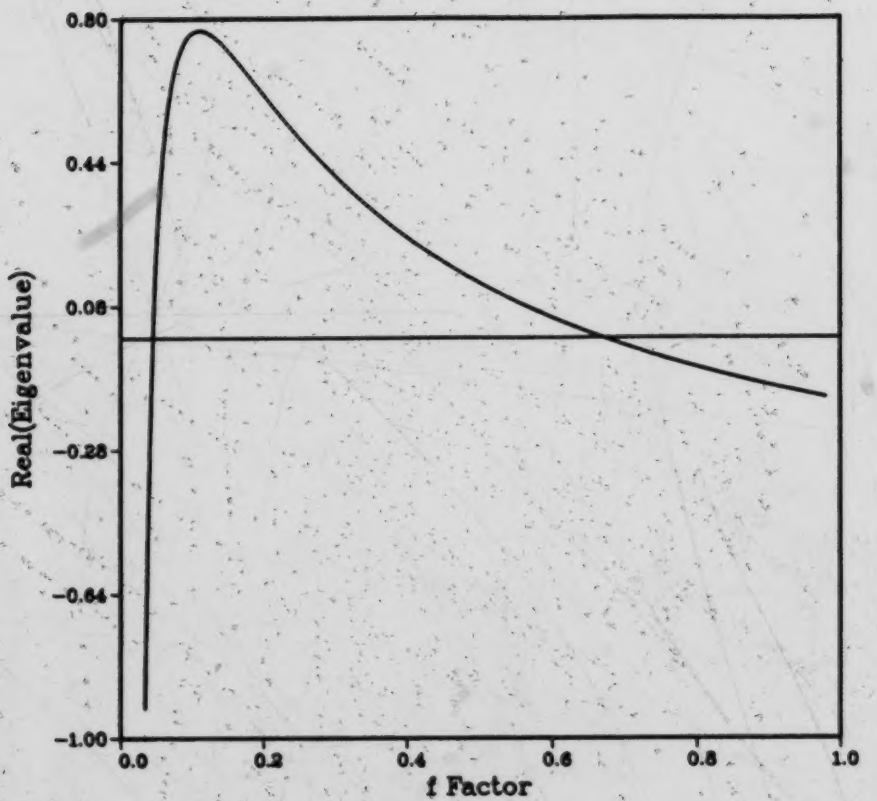


Figure 1.10 shows the relation of the real part of eigenvalue versus the factor *f*. The sharp increase of the eigenvalue with *f* indicates a sudden transition of the system from stable to unstable.

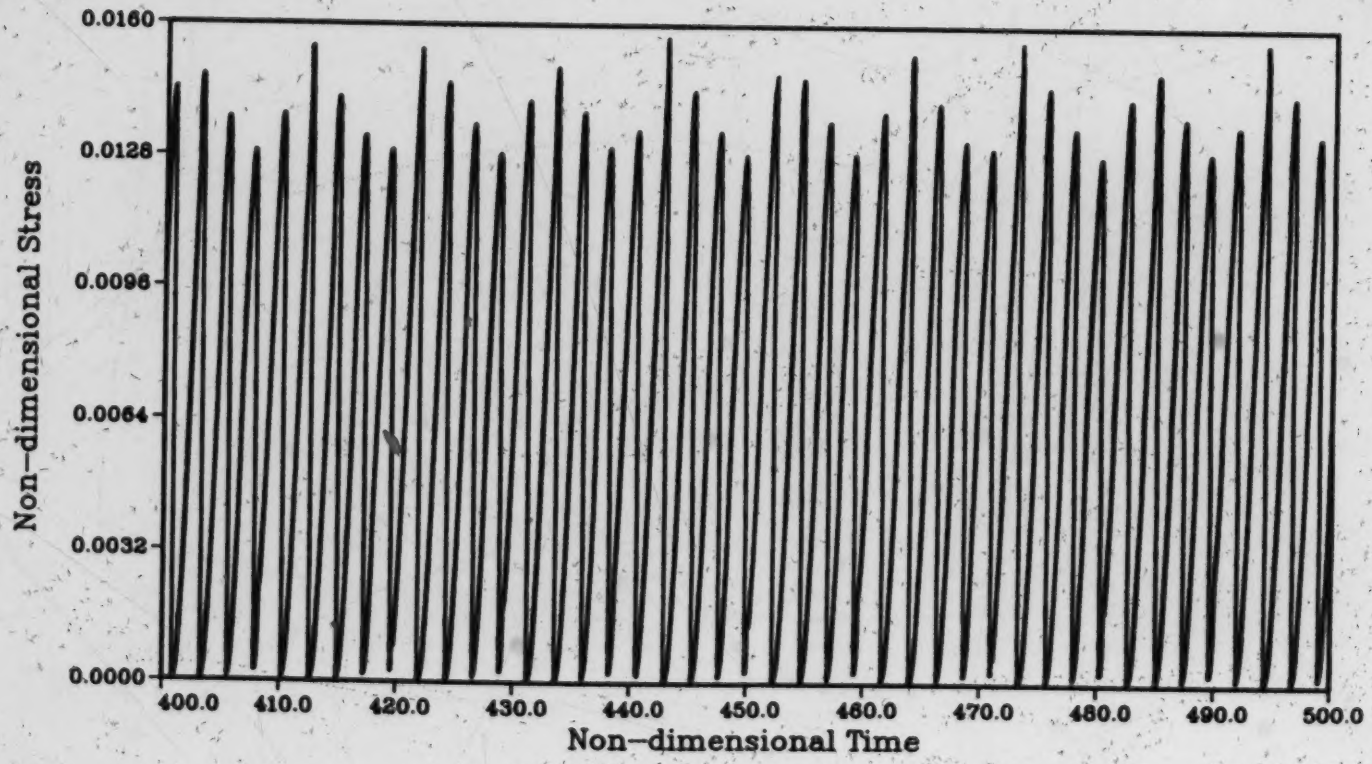


Figure 1.11 shows the stress of one fault segment versus time for the model of two fault segments.

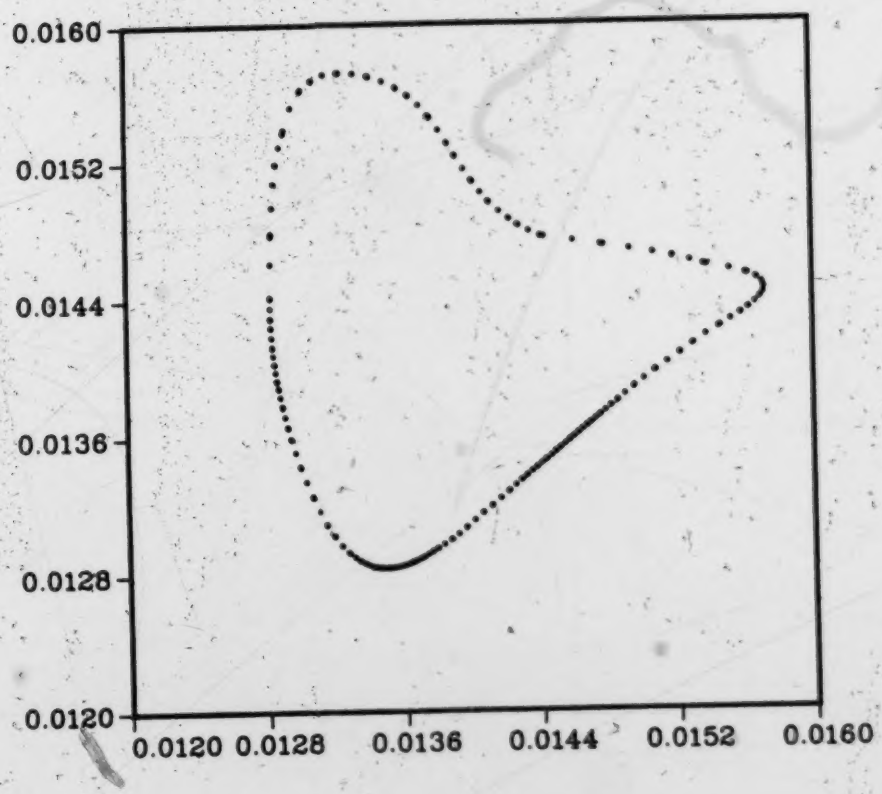


Figure 1.12 shows the one-dimensional map of successive maximum stress.

Materials	E ($\times 10^{12}$ dyn/cm^2)	ν	ρ	Faults	E ($\times 10^5$ dyn/cm^3)	ν
1	1.2	0.25	3.3			
2	1.04	0.3	2.295	1	2.0	0.4
3	0.89	0.25	2.92			
4	1.163	0.3	3.3			
5	0.71	0.3	2.92			
6	1.25	0.3	3.285			

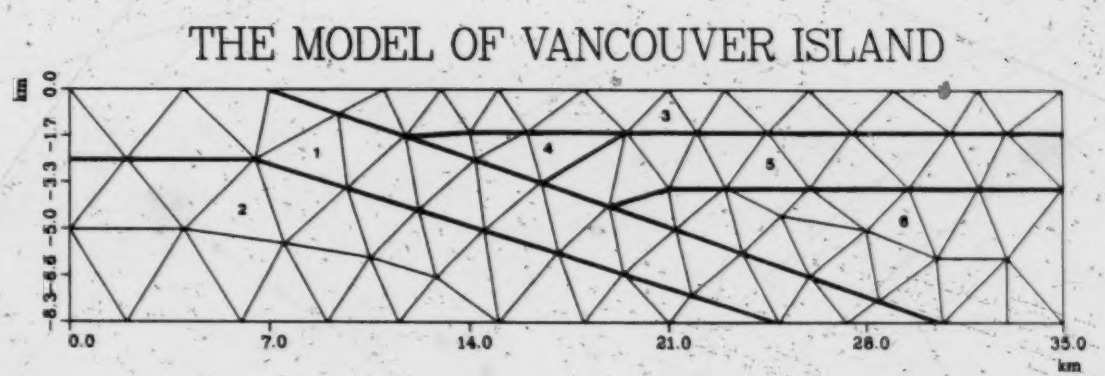


Figure 1.13 shows the finite element mesh of the two dimensional model of the west coast of Canada. The above table shows the elastic constants of each material.

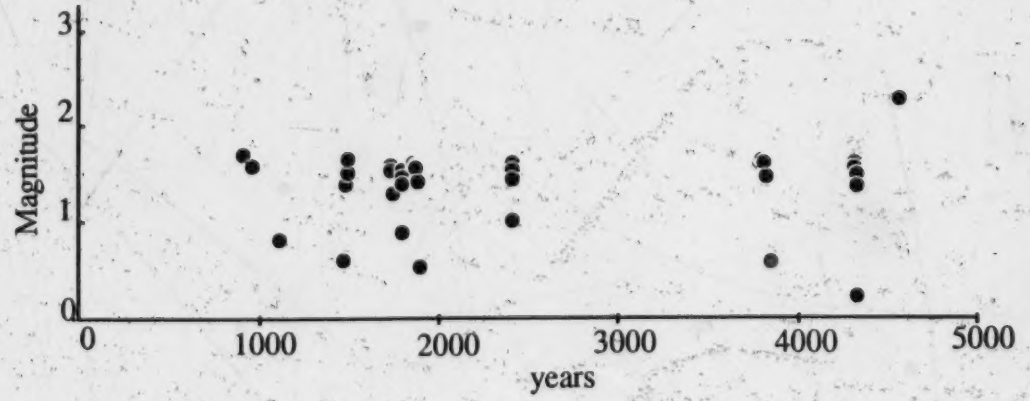


Figure 1.14 shows the magnitude of simulated earthquakes versus time.

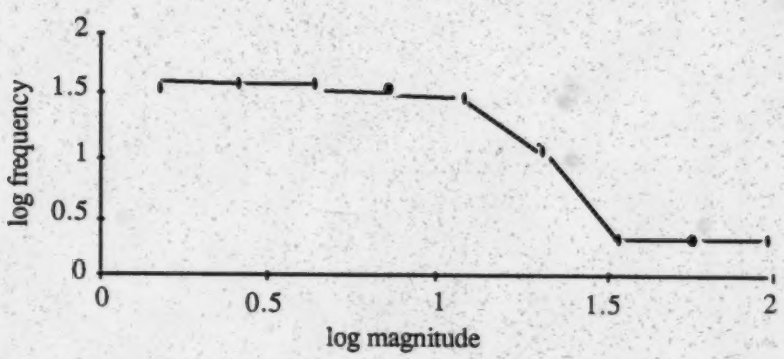


Figure 1.15 shows the magitude frequency relation for the simulated earthquakes.

Chapter 2: Is the Dynamics of the Lithosphere Chaotic ?

Lithosphere dynamics can be represented by a stochastic model (Knopoff, 1971; Kagan and Knopoff, 1987; Kagan, 1988) or by a deterministic model that usually considers only a single fault in isolation from the others (Rudnicki, 1988) and emphasizes the onset of instability of the model. Long term behavior of seismicity seems to reflect a stochastic process but it is not necessarily true that complex behavior must be due to complicated causes. Recent development in non-linear dynamic theory suggests that the random behavior of a system can be inherent to the system, rather than imposed by the environment or other uncontrolled forces. Clearly simple systems also behave randomly. This inspired several studies into the possibility of chaotic earthquake occurrence (Huang and Turcotte, 1990; Carlson and Langer, 1989). The results from the modelling discussed in last chapter and those by others are encouraging. It may be possible to extend such a view of the fault models to a general view of the dynamics of the lithosphere, and to treat the dynamics of the lithosphere as a non-linear system.

In the next few sections, I will explore the dynamics of the lithosphere in terms of the idea of chaos. This suggests that the lithosphere dynamics is better described by chaos rather than by randomness. Such a view of lithosphere dynamics suggests that problems such as earthquake prediction should be addressed in a completely different way, and that some useful information about the general behavior of a non-linear dynamic system can be used in the study of earthquake occurrence. The following discussions are heuristic rather than mathematically rigorous, the latter is impossible now in the study of lithosphere dynamics. Our understanding of the lithosphere dynamics is still in its infancy. The following discussion attempts to bring the lithosphere dynamics closer to the realm of quantitative science.

2.1 Chaos Theory and Its Implications on Lithosphere Dynamics

If an explicit integration of a group of differential equations is impossible, one

increases emphasis on qualitative behavior of physical systems, and examines geometric properties of phase portrait in phase space where its evolution can be visualized. For example, the motion of a simple pendulum can be expressed in a phase space with the position and velocity of the pendulum as coordinates. In the presence of friction, the phase curve is a spiral which approaches a fixed point known as an attractor since it attracts nearby orbits. Starting from any point in phase space, the system will come to rest with a passage of time. Therefore, knowing the property of the fixed point, one will know the asymptotic behavior of the system (Crutchfield et al., 1986). More generally, instead of a fixed point phase curves can be attracted to a smaller region, a limit cycle for instance. Therefore, an attractor in general is a lower dimensional object embedded in a phase space whose dimensionality is associated with the number of a priori degrees of freedom of the system. When the state of the system is close to an attractor, the number of degrees of freedom of the system may be greatly reduced.

Most dissipative systems are characterized by an attractor. This makes things much easier since we don't need to consider the rest of the phase space, only the attractor. However, an attractor can become very "strange" when it does not attract in all directions. It may stretch in some directions while contracts in others. The volume element of the attractor will be stretched and get folded at same time to remain confined in a bounded domain. A simple example to illustrate this operation is the usual baker's transformation (Schuster, 1988),

$$x_{n+1} = 2x_n \bmod 1$$

$$y_n = \begin{cases} a\,y_n & \text{for } 0 \le x_n < 1/2 \\ 1/2 + a\,y_n & \text{for } 1/2 \le x_n \le 1 \end{cases}$$

When $a = 1/2$ the map is area preserving, and the map is non-area preserving when $a < 1/2$. The transform stretches a unit square in the x direction and contracts it in the y direction. After the unit square be stretched it is then folded back to keep the object within the unit square. The object is made a strange attractor after repeated application of this map. Figure 2.1 illustrate the operation of the baker's transformation on an object within an unit square. After a few steps, the object becomes unrecognizable.

an unit square. After a few steps, the object becomes unrecognizable.

Chaos mixes orbits in phase space ways like the baker's transformation. Nearby trajectories on a chaotic attractor will stretch first, and then fold back on themself. After some time the attractor will have a multisheeted structure, and become cantor-set like in some directions, and accordingly a fractal (Mandelbrot, 1983). Such a complex orbits therefore lead to a random behavior of the system.

Another property of a chaotic attractor is that any error associated with the initial condition will increase exponentially. For a physical system, it is impossible to make precise measurements without any error. Therefore the state of the system is within a small phase volume. Such a small volume stretches and folds due to the chaotic behavior of the system. We will soon loss the track of the state of the system. Any microscopic fluctuations in the system will be amplified to macroscopic scale on the chaotic attractor, and therefore causes loss of the ability to predict. The prediction of future made for such kind of system is effective only in short term.

The idea of strange attractors entered physics with Lorenz's remarkable discovery of a fundamental regularity in a chaotic system (Lorenz, 1963). Lorenz equations are a severe truncation of the Navier-Stokes equation, motivated by the desire to understand the predictability of the weather. His system has three ordinary differential equations:

$$\dot{x} = -\sigma x + \sigma y$$

$$\dot{y} = -xz + \gamma x - y$$

$$\dot{z} = xy - bz$$

Nevertheless, the system behaved in an apparently random fashion. It is the first example of a low-dimensional system that displayed a complex behavior.

Such systems are deterministic, but the output data sequence from the system can be very chaotic. The apparent randomness shown in the output is inherent to the system; it cannot be removed by a better understanding of any part or aspect of the system. A chaotic data sequence may behave statistically the same as those generated by a stochastic process. It may pass the conventional statistical tests of randomness. Some simple maps, a triangular map for instance, can produce uncorrelated time series, ie. $\langle x_t\, x_{t+j} \rangle = 0$ unless j

random phenomena. If a system has a wide distribution of characteristic energies it should be described as a stochastic process. Other may relate to chaotic behavior generated by a deterministic system of a few degree of freedom. The possibility of the latter provides us an alternative view of the dynamics of the lithosphere that appears quite random. We hope that we may possibly find a deterministic model to describe the dynamics of the lithosphere.

Occurrence of earthquakes is an inseparable part of the dynamics of the lithosphere. Therefore, the understanding of earthquake occurrence is a key to the understanding of the dynamics of the lithosphere. Data about mechanical state of the lithosphere from direct underground measurement is sparse and poor. This situation will not improve soon due to the tremendous difficulties involved in the in situ experiments required to gather such information. Earthquakes, however, provide an essential cognitive link with the dynamics of the lithosphere, and the main way we can understand it is through interpretation of earthquake occurrences. It should come as no surprise that earthquakes become important phenomena, exemplifying some aspect of our views on the evolution of the dynamics of the lithosphere.

The statistical analysis of the nature of earthquake occurrence shows that the dependence between events is very weak (Keilis-Borok, et al., 1971). Several other statistical investigations led to the conclusion that the mainshocks are mainly Poissonian distributed (Gardner and Knopoff, 1974). Those statistical results suggest a complex lithosphere dynamics but recent advances in non-linear dynamics have shaken such a view of the dynamics of the lithosphere. Further investigation by Kagan and Knopoff (Kagan and Knopoff, 1980; Kagan, 1981a,b) show that earthquake fault zones have a three dimensional structure that has a fractal distribution of size. The fractality implies that the underlying dynamic field governing the occurrence of earthquakes does not have an intrinsic length scale. This suggests that an analogy to structures associated with other non-linear dynamic systems, which are deterministic, may exist in the earth. The dynamics of the lithosphere may be non-linear and deterministic.

2.2 Reconstruction of the Lithosphere Dynamics from Earthquake Catalog

The model discussed in chapter 1 is simple, but very instructive. In that simplified model the state of the fault can be described by a few state parameters, such as stress and temperature. A group of differential equations describes the evolution of the dynamics and the onset of an earthquake. I assume this idea can be extended. The state of the lithosphere may be described by a few state variables. The lithosphere dynamics is described by a group of partial differential equations.

Such a system described by a group of partial differential equations usually has an infinite number of degrees of freedom. A simple way to reduce a partial differential equation to a group of ordinary differential equations is to truncate its series expansion in the way Lorenz truncated the non-linear Navier-Stokes equations for the atmosphere. The procedure is necessary since the current non-linear dynamic theory only applies to systems described by a group of ODEs. A simple and straight forward way of illustrating the discretization of the dynamics of the lithosphere into a group of ODEs is to follow the exercise in chapter 1, where the fault along the subduction plate is coarse graining into 10 small segments. Each segment depends on two ODEs that couple to the ODEs for other segments.

We can, in principle, coarse grain a region of lithosphere into N small patches (Figure 2.2a). The state variables in each patch can be treated as uniform. If each patch has k state variables, the system has kN coupled ODEs. The state of the system at a given time will correspond to a point in a kN dimensional phase space. The evolution of the system will draw a trajectory in this space. If there exists an attractor, and one exists for most dissipative systems, the trajectory will settle on it after the transient period.

To relate individual seismic events to the evolution of the dynamics of the lithosphere we need to introduce the idea of a Poincare map. The abstract idea often has value for discretization of a continuous system. To this end consider a piece of hypersurface Σ which is one dimension less than phase space. We then consider the successive intersection of the trajectory with Σ, and obtain a series of points on the hypersurface. The time evolution of the governing differential equation is then reduced to a

series of successive points on this surface.

It is often easier to study such a discrete mapping than a differential system. Many ideas that are somewhat cumbersome to state for ordinary differential equations are more clearly seen for a Poincare map because the irrelevant details of the short term evolution vanish there. Experience shows that the essential properties of the differential system reflect in equivalent properties of the mapping (Henon, 1981). For example, a simple periodic orbit of the differential system, closing back upon itself after one revolution, corresponds to a fixed point of the mapping. The periodic orbit is stable if and only if the fixed point is stable (Wiggins, 1988). We can thus forget about the system of differential equations and consider mapping instead.

Here, we can imagine selecting a special surface of section in a kN dimensional phase space. The lithosphere states on the surface correspond to the states when an earthquake will occur. Each time the trajectory crosses the surface corresponds to the moment of on-set of an earthquake. The time needed for the trajectory to leave the surface and return, will correspond to the interval between successive events (Figure 2.2b). A collection of the points marked on the surface will give us an earthquake catalog. We can, therefore, consider the earthquake catalog only and ignore the rest of trajectories, ie. the detail of what happens between intersections. Here we may disregard the physics and view the system as a black box which produces a stream of seismic events. We explore then how the geometric properties of the earthquake catalogs reflect the equivalent properties of the dynamics of the lithosphere. All our questions about the lithosphere dynamics are posed in terms of these outputs.

Such an interpretation of earthquake catalogs seems artificial, but it helps understand the role that the catalogs play in the dynamics of the lithosphere. The advantage of this consideration is that the earthquake catalog is now viewed as an output time series of a deterministic system, and the data analysis techniques developed recently for studying such systems are applicable. The justification for selecting such a surface of section is this: Each earthquake will correspond to a point in kN dimensional phase space, and if the dimension of such a collection of points is one or more dimension less than the embedding space, ie. less than kN, selecting such a surface of section should be possible.

All this suggests that if the surface of section exists, and if the existence and uniqueness theorem holds for all times for the differential equations governing the lithosphere dynamics, the solution of the equations can be written out as an vector $s_n(t)$, where $t_n < t < t_{n+1}$, for the initial condition s_n. This guarantees the existence of mapping G:

$$s_{n+1}=G(s_n)$$

The problem of earthquake prediction then becomes the study of the mapping G. If we know a state of the lithosphere when a certain earthquake occurred, the state of the lithosphere at which next earthquake will occur can be determined by G. In this case, there will be no increase in predictive ability upon learning more historic states s_{n-1}, s_{n-2} ··· However, when the measurement is incomplete, a subset of s_n for example, it can only partly constrain the system state. The degree of freedom left out can be determined by considering more history measurements.

One may notice that the earthquake catalog mentioned above differs from real earthquake catalogs that are obviously an incomplete measure of the state of the lithosphere. The real catalog contains only occurrence time, epicenter and magnitude of each event but no other information associated with the state of the lithosphere, such as stress and temperature. They are definitely important to the onset of a seismic event. Can we hope to understand the system without knowing all this information?

This is possible because we restrict our attention to the dynamics of a finite dimensional attractor, and therefore the information can be retrieved from the historical data. Packard, et al. (1980) and Takens (1981) have proved the existence of an embedding from an m-dimensional manifold to a n-dimensional Euclidean space R^n defined by $\Phi(x)=(\nu(x), \nu(\phi_1(x)),..., \nu(\phi_n(x)))$ when $n \geq 2m-1$, where ϕ_t is the flow of the system, and ν a smooth function on the manifold. This forms the basis of reconstruction techniques for phase portraits from time series measurements in experimental domains.

In practice it is necessary to relate this embedding theorem to a time series of measurements made on the system. A method of reconstructing a phase portrait from a

time series is to create more signals from a single one by using time delay (Grassberger and Procaccia, 1983; Roux, at el., 1983). Consider a time series {B(t)} and time delay T. Suppose we create a state vector x(t) by assigning coordinates $\{x_1, x_2, ...x_n\}$ by { B(t), B(t+T), ..., B(t+(n-1)T) }. If the dynamics takes place on an attractor of dimension D, then a necessary condition for determinism is $n \geq D$. Besides, an n-dimensional phase portrait constructed from the vectors { B(t), B(t+T), ..., B(t+(n-1)T) } will have the same properties (topologically) as one constructed from measurements of m independent variables, if $n \geq 2m+1$, where m is the dimension of a manifold containing the attractor.

In the case of the dynamics of the lithosphere, the portrait of the phase space can therefore be reconstructed from the earthquake catalogs. With the above discussion, we arrive at following conclusions: The determinism of the lithosphere dynamics is ensured if we use sufficient number of historic events in the earthquake catalog, and this leads to the possibility of predicting future earthquake by decoding the earthquake catalog. If premonitory seismicity patterns associated with nature of the attractor exist at all, they can be found in earthquake catalog as long as we use sufficient number of historic events, since the phase portrait obtained in this way preserves its geometrical invariants of the dynamics. However, such arguments are effective and useful in practice only if the attractor is characterized with a low dimensionality since, until present, only low dimensional attractors have been successfully reconstructed in the experimental domain (Roux, et al., 1983).

2.3 Measuring chaos in lithosphere dynamics

Any mechanical system on the scale of classical mechanics is deterministic. Randomness is only a matter of degree. It occurs to the extent that something cannot be predicted, and that usually depends on the available information. Such unpredictability depends very much on the dimensionality of an attractor. If a time series is produced by motion on a low dimensional attractor, then the motion is not random and it is possible for us to make predictions. If the dimension of the attractor is large enough, the amount of data needed to make a good prediction may be prohibitive. The problem gets exponentially

worse as the dimension increases (Froehling, et al., 1981). The dimensionality therefore can be used as an important criterion in distinguishing systems associated with low dimensional attractors from those with high dimensional ones.

The dimension of an object describes how its volume scales with increasing linear size. Various dimensions have been defined for this purpose. Of these Hausdorff dimension is most commonly used. The Hausdorff dimension, also referred as fractal dimension by Mandelbrot (1983) for a self-similar set is given by (For a general definition see Farmer (1983) and Eckmann (1985))

$$D_h = \lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{\log(1/\varepsilon)}$$

where, if the set in question is a bounded subset in m-dimensional Euclidean space, then $N(\varepsilon)$ is the minimum number of m-dimensional spheres of diameter ε needed to cover the set. To illustrate the idea of the dimension for a self-similar object, first consider a square. Since a plane has Euclidean dimension of 2, it follows that a unit square can be divided into $N=b^2$ squares of side $r=1/b$, b is number of subintervals divided on the side of the square, and the length of the subintervals is $r=1/b$. The dimension in this case is given by $d=\log N/\log(1/r)=2$ (Figure 2.3a). The pattern shown in Figure 2.3b is a Sierpinski carpet (Mandelbrot, 1983). It can also be decomposed into reduced-size pieces, and with $r=1/3$, we have $N=8$. The resulting dimension in this case is $d=\log N/\log(1/r)=1.893$.

Fractal dimension is a purely geometrical measure, independent of the frequency with which a trajectory visits the various parts of the attractor. Another dimension which takes into account the frequency of visiting, is the information dimension. Using the partition of phase space into cells with diameter ε, the information, which measures the average knowledge gained after we have learned that the system is in a specific state, can be written as:

$$I(\varepsilon) = -\sum_{i=1}^{N(\varepsilon)} P_i \log P_i$$

where P_i is the probability for the trajectory to fall into the i-th cell. The information

dimension is defined as

$$D_I = \lim_{\varepsilon \to 0} \frac{I(\varepsilon)}{\log(1/\varepsilon)}$$

Note that $I(\varepsilon)=\log N(\varepsilon)$, if all cells have equal probability, and hence $D_h = D_I$. However, for the general case the information dimension always a lower bound to the Hausdorff dimension (Grassberger and Procaccia, 1983).

Consider now (Schuster, 1988) an infinite set of dimensions D_0, D_1, D_2, ..., which relates to the f-th power of P_i via

$$D_f = \lim_{\varepsilon \to 0} \frac{1}{f-1} \frac{\log\left(\sum_{i=1}^{N(\varepsilon)} P_i^f\right)}{\log \varepsilon} ; \qquad f = 0, 1, 2, ...$$

where P_i is defined as before. When f = 0 and 1, we obtain the Hausdorff dimension and information dimension respectively. Numerical determination of the dimension D_f by covering the phase space with a set of cells of diameter ε and counting the number of iterates which lie in a certain cell is rather cumbersome and, in fact, impossible for high dimensional attractors. However, things become simpler when f = 2. In this case

$$\sum_{i=1}^{N(\varepsilon)} P_i^2 = \text{the probability that two points of the attractor lie within a cell}$$

$$= \text{the probability that two points at the attractor are separated by a distance smaller than } \varepsilon$$

$$= \lim_{N \to \infty} \frac{1}{N^2} \{\text{number of pairs (i,j) whose distance } |s_i - s_j| \text{ is less than } \varepsilon\}$$

where s_i and s_j denote the states of the system at time t_i and t_j. This quantity is defined as correlation integral, denoted by $C(\varepsilon)$. It measures the spatial correlation of the points that lie on an attractor. Most pairs will be dynamically uncorrelated due to the exponential divergence of trajectories, but the spatial correlation will remain if the trajectories are confined to an attractor whose dimension is lower than the phase space. Similarly, D_2,

which is called correlation dimension, is given by:

$$D_2 = \lim_{\varepsilon \to 0} \frac{\log C(\varepsilon)}{\log(\varepsilon)}$$

Numerical calculation of D_2 is relatively easy and straight forward; it is thus a most widely used parameter for characterizing an attractor. Generally we have following relation:

$$D_{f'} \leq D_f \qquad \text{for } f' > f$$

where the equal sign holds only when the attractor is uniform (Schuster, 1988). Of these dimensions, the correlation dimension, therefore, gives a lower bound to all the others. According to the discussion in the last section, the correlation integral $C(\varepsilon)$ can be calculated from a single experimental measurement by applying the method of delays to the time series. The method has brought a transition from purely mathematical and theoretical results to the quantitative determination of chaotic effects in experimental data. The correlation dimension D_2 therefore becomes an important parameter to indicate the randomness of a physical system.

The dimensions are defined as a local property of a fractal set. In a purely deterministic system, the complete past and future history of the dynamics is determined by the infinitesimal structure to be found at any part of the attractor. Thus we expect that a chaotic attractor has locally the same dimension everywhere. But for a real system the base length scale may limit the degree of structure the actual attractor can have. So it is expected that the dimension will usually vary across the attractor (Shaw, 1985). The correlation dimension D_2 calculated from a time series in the way discussed above should therefore yield an average value for the dimension of the attractor.

The calculation of the correlation dimension D_2 was carried out for the west coast of Canada, from 118°W to 130°W and from 45°N to 51°N, which is a tectonically active area. The earthquake catalog for this area, selected from the seismicity catalog of the Geological Survey of Canada and the U.S. National Earthquake Information Service, has a sufficient number of small earthquakes since the early 1950's. (See more discussion on the

catalog in chapter 4). In the period of 1950 to 1985, the catalog contains 5200 seismic events, among them there are 3160 events with magnitude above 2; 360 mainshocks with magnitude above 4 and 2000 mainshocks with magnitude above 2. The mainshock catalogs are generated after the aftershocks are deliberately removed according to a certain empirical time and space windows because they are the events closely related to mainshocks. The calculations were carried out for both mainshock catalogs and the catalog including aftershocks in order to compare the results respectively, and see how aftershocks affect the integral dynamic evolution of the lithosphere.

The set of time intervals between successive events are used as the data series for the calculation, because they are accurately recorded and their physical meaning is obvious. Figure 2.4 shows the result of the calculation for mainshocks of magnitude above 4. Successive curves are generated by taking the data in successive groups of 3, 6, 9 ..... and so on. Each curve corresponds to an increase of the dimension of the phase space by three. The observed dimensionality of the attractor should also increase until the dimensionality of the embedding space is large enough for the attractor. Once there is enough room in the embedding space, larger than 2m+1 according to Takens (1981), the dimension will then arrive at a constant value m, ie. the dimension of the attractor. This dimension is measured from the slope of the curves in Figure 2.4; it is between 5 and 6. The same calculation has also been done for the mainshocks of magnitude above 2. Figure 2.5 displays the result of the calculation. The slope of the curves is between 11 and 12. There is no clear indication that the slope reaches saturation. Figure 2.6 shows the result of the calculation for the catalog containing the events of magnitude above 2, including aftershocks. The slopes of the curves saturate at a value between 3 and 4.

There are a few possibilities that can cause the inconsistency of the results from the calculations for the catalogs which contain mainshocks whose magnitudes are above 4 and above 2. One is that errors can be introduced into the calculation due to incomplete recording of small events. Since the data used are time intervals between successive events, missing events could greatly distort the data, and therefore make the data appear more noisy. We know that the dimension for a random noise is infinity. More noise can also be brought into the mainshock catalogs by the operation of removing aftershocks from

the catalog. The effects become significant for the small mainshock events whose epicenters generally have a large uncertainty. It will therefore affect the selection of space window for removing aftershocks.

The low dimensionality shown in the catalog which contains all the events indicates that the aftershocks are integral part of the whole dynamics. It contains significant amount of information about the dynamics of the lithosphere. Removing aftershocks from the catalog will eliminate useful information about the lithosphere dynamics. Certainly there is another possibility that aftershock events are characterized by a smaller length scale in the attractor than that of mainshocks since they are local events of mainshocks both in time and space. The section of the attractor corresponding to aftershocks has a lower dimension than that of mainshocks. The overall effect of the dimension across the attractor is less than what we obtained from the mainshock catalog.

The error in the correlation integral caused by the finite length of data series is relatively small. It can be shown that for a data series with n data, the error will less than 1/n (Appendix B). Therefore, this error is not significant in the above calculation. The main source of error in this case probably come from missing events in the catalog. Since the lithosphere is a spatially open system, when we use a space window to select a study area, some events which are significant to the dynamics may be left outside of the window. It is not clear, however how those missing events will affect the reconstruction of the phase portrait.

Another source of error is due to the way of investigating the scaling of the correlation function. Ruelle (1990) showed that the correlation dimension calculated from a data series of length N is meaningful unless it satisfies (Appendix B):

$$\text{correlation dimension} \leq 2\log_a N$$

He argued that the measurement should take at least over one decade, therefore, a should be 10. Only those dimensions which are well below $2\log_{10}N$ have credibility. In the above exercise only the analysis for the catalog containing all the events meets this requirement. The analysis for the catalog contains mainshocks of magnitudes above 4 is far from it. The region within which the correlation function behaves linearly in the log-log

plot is only across a scale of 3 for the mainshock catalog with magnitude above 4. Therefore the results obtained from the analysis are far from conclusive.

2.4 Application of singular spectrum analysis to earthquake catalog

The method proposed by Grassberger and Procaccia (1983) for computing the correlation dimension of strange attractor is powerful,but its validity for a short and noisy data series is questionable (Ruelle, 1989), especially when data is polluted with white noise. Earthquake catalogs are data series of this kind; short and noisy. Nevertheless, with such a short and noisy records there may be some quantities other than the correlation dimension that one may try to compute; they may be better behaved and still able to indicate the underlying order in the data.

The answer to this question is provided by the singular spectrum analysis proposed by Broomhead and King (1986). It is a sophisticated linear analysis of a time series which was intended to provide good phase reconstructions. It is an application of the Karhunen-Loeve expansion for random processes (Fukunaga, 1972). The applications of the method to the paleoclimatic time series, which contain only a few hundred data samples, and to the earthquake catalog from the Parkfield, California showed significant results (Vautard and Ghil, 1989; Horowitz, 1989).

The basic idea of singular spectrum analysis (SSA) is to apply an n-window, which makes visible n elements of the time series, to obtain a sequence of vectors in R^n by using the time delay method. These vectors are then used to form the rows of a trajectory matrix, X, which will therefore contain the complete record of the patterns that occurred within the window. If the rank of the matrix X is lower than n, ie. the number of the linearly independent vectors is less than n, the patterns within the window are only a subset of all n-element patterns in R^n. The rank will indicate the dimensionality of the subspace in which the embedded manifold is to be found. This dimension provides a reasonable upper bound to the dimension of the attractor m, but it is not an invariant of the embedding process.

Broomhead and King showed that the rank of the matrix X can be found by

examining the eigenvalues of the structure matrix $\Theta = XX^T$. The analysis can be done equivalently with the covariance matrix $\Xi = X^TX$, since the rank of the matrix Θ is same as the rank of Ξ. The matrix Ξ gives the time averaged correlation between all pairs of elements in the n-window, and it is more tractable than Θ since it is implicit to the approach that the embedding dimension is small.

When the matrix is polluted with experimental noise, Broomhead and King showed that every singular value, the eigenvalues of Ξ will be augmented by the variance of the noise. Thus noise dominates any of those components whose corresponding eigenvalues are comparable to the variance of the noise, and such components should be discarded. The significant singular values which are above the noise level, represent the deterministic aspects of the time series and the remaining singular values represent noise.

The SSA was applied to the earthquake catalogs discussed in last section. The spectra shown in figure 2.7 are calculated when the n-window is chosen 20. The spectra for the catalogs containing mainshocks of magnitude above 4 and 2 and the catalog containing all the events of magnitude above 2 are displayed in figure 2.7a, 2.7b, and 2.7c respectively. They do not show the simple structure proposed by Broomhead and King in which the singular values decrease monotonically until they merge into the flat noise floor, which obtains only when the noise is truly white. Several plateaus occur in each spectrum. Surprisingly in all three cases, there are three singular values above the first plateau which are probably associated with a noise floor. This agreement strongly supports the result from the correlation dimension analysis for the catalog containing all the events in last section. It supports the argument that aftershocks contain significant amount of information about the integral dynamics of the lithosphere.

2.5 Discussion

What we are really interested in at the present stage of the study on the dynamics of the lithosphere is whether we should view the system as deterministic or stochastic. The answer to this question depends on the dimension of the attractor associated with the

lithosphere dynamics. If the dimension obtained for an experimental signal is less than F (the dimension of phase space) one knows that signal stems from deterministic chaos rather than random noise, since random noise will always tend to fill the phase space and result in a dimension F. The results of the analysis in above sections are encouraging. Both the analysis of the correlation dimension and the singular spectrum analysis suggest the existence of a low dimensional attractor. However, it is by no means conclusive.

It is necessary to point that distinguishing a deterministic process from a stochastic one according to the value of the correlation dimension was questioned recently by Osborne and Provenzale (1989). They demonstrated that a stochastic self-affine time series can also yield a finite correlation dimension, and argued that such a correlation dimension corresponds to the fractal dimension of the fractional Brownian motion trail. Note that all methods for experimental determination of the attractor dimension rely on a certain space scale. If the fractality is generated by deterministic chaotic dynamics then small space scales are associated with long time scales (close returns on the attractor). On the other hand the small space scales associated with a Brownian motion trail are related to small time scales. Therefore we can distinguish them according to this nature of the time series. A fractional Brownian motion, $V_H(t)$, has a variance (Voss, 1989; Mandelbrot, 1985)

$$\langle V_H^2(t) \rangle \propto t^{2H}$$

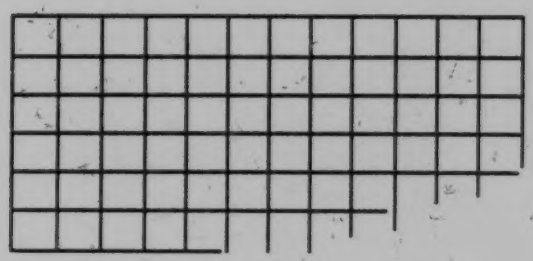
when we choose $V_H(0) = 0$. It is obvious that $V_H(t)$ is not a stationary process, while a time series generated by a chaotic dynamics is generally stationary. If we scale $V_H(t)$ by the factor t^{2H} to make $V_H(t)$ stationary, the scaled time series will have a exponent H=0. As demonstrated by Osborne and Provenzale such time series have a dimension of infinity.

I would like to emphasize again the conceptual framework discussed in section 2.2, which is based on the embedding theorem proposed by Packard and Takens. It is important to notice that the phase portrait reconstruction from an earthquake catalog conceptually links the earthquake catalog to the entire dynamics of the lithosphere. This broadens our understanding of the earthquake catalog. It is possible for us to link, in the sense of topological equivalence, the earthquake catalog to other geophysical observables. Some of them, strain meter measurements or the earth tilt measurements for example,

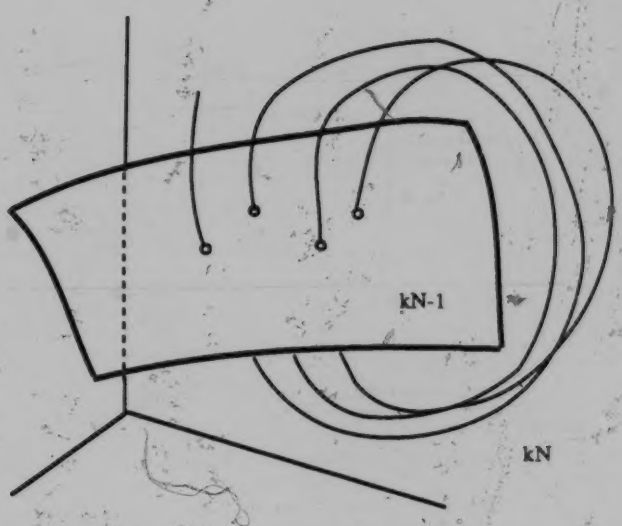
seemingly unrelated to earthquake occurrence may also be used for the study of earthquake prediction.

The existence of a low dimensional attractor in the lithosphere dynamics also has its implications on the way of analysing the dynamics of the lithosphere and the strategy in selecting what kind of model to describe the system. Successful physical models generally have small number of degrees of freedom. If it is true that the dynamics of the lithosphere is deterministic, ie. it characterized by a low dimensional attractor, we may possibly construct a deterministic model of a few degrees of freedom to describe the dynamics of the lithosphere.

Figure 2.1 illustrates idea of baker's transformation. An object is stretched in one direction while contracted in others and then gets folded at same time in order to remain confined to a bounded domain.

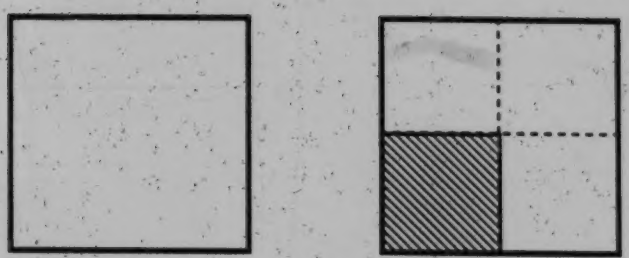


(a)

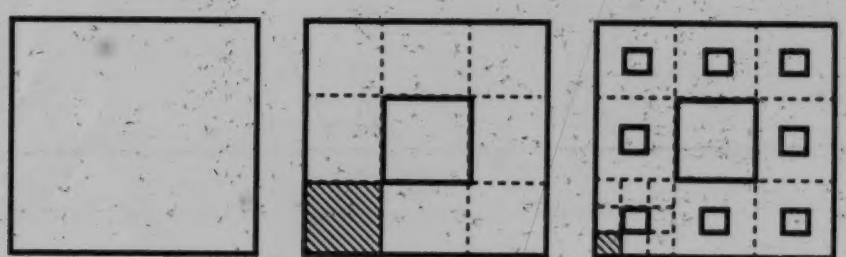


(a)

Figure 2.2 (a) illustrates the idea of discretization of the lithosphere into N small patches and each patch is described by k state variables. (b) illustrates the idea of considering earthquake catalogs as corresponding to Poincare points.



(a)



(b)

Figure 2.3 illustrates the idea of fractal dimension. (a) shows a plane that has a dimension of 2. (b) shows a Sierpinski carpet that has a dimension of 1.893.

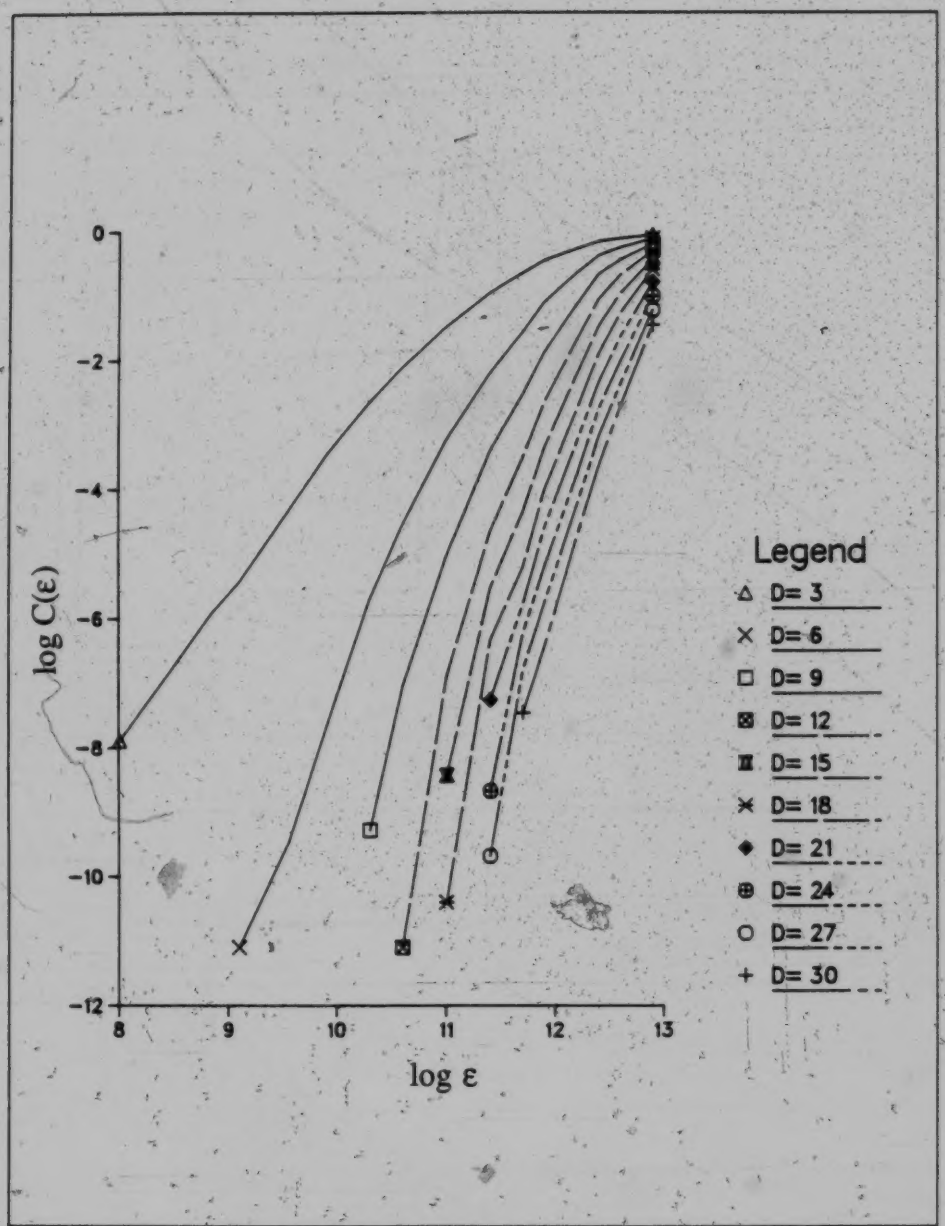


Figure 2.4 shows logC(ε) versus log ε for different D calculated for the mainshock catalog containing earthquakes of magnitude above 4.

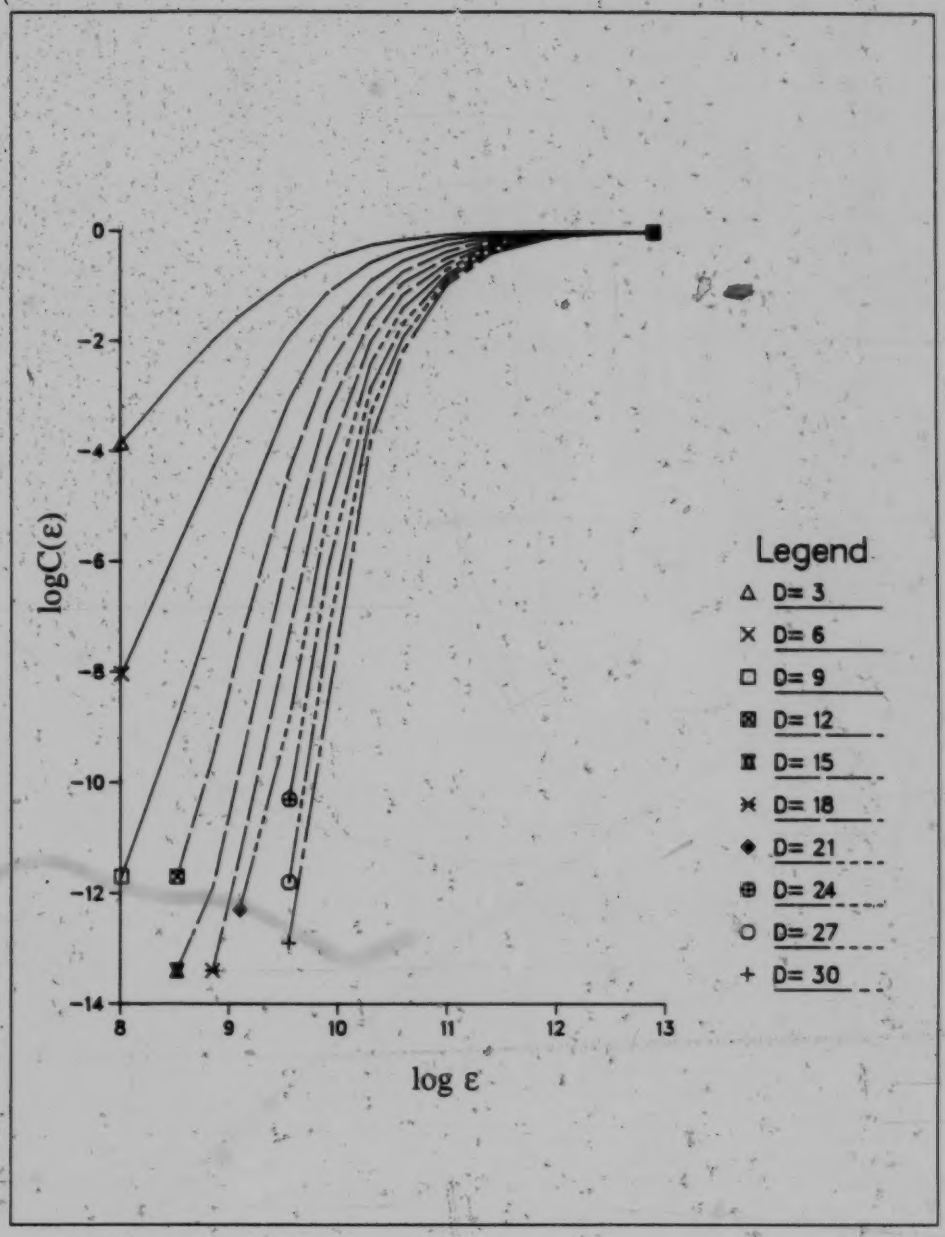


Figure 2.5 shows logC(ε) versus log ε for different D calculated for the mainshock catalog containing earthquakes of magnitude above 2.

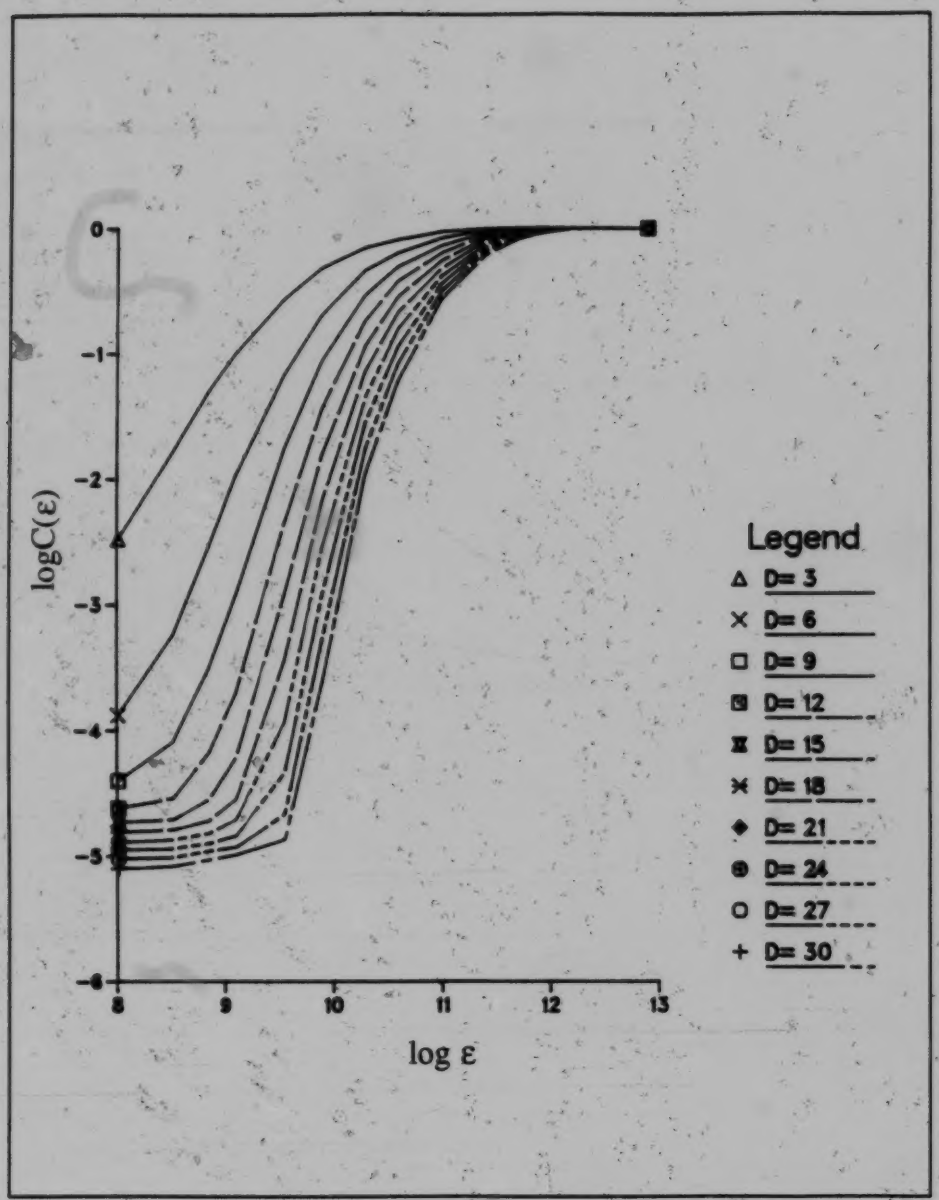


Figure 2.6 shows logC(ε) versus log ε for different D calculated for the catalog containing all events of magnitude above 2.

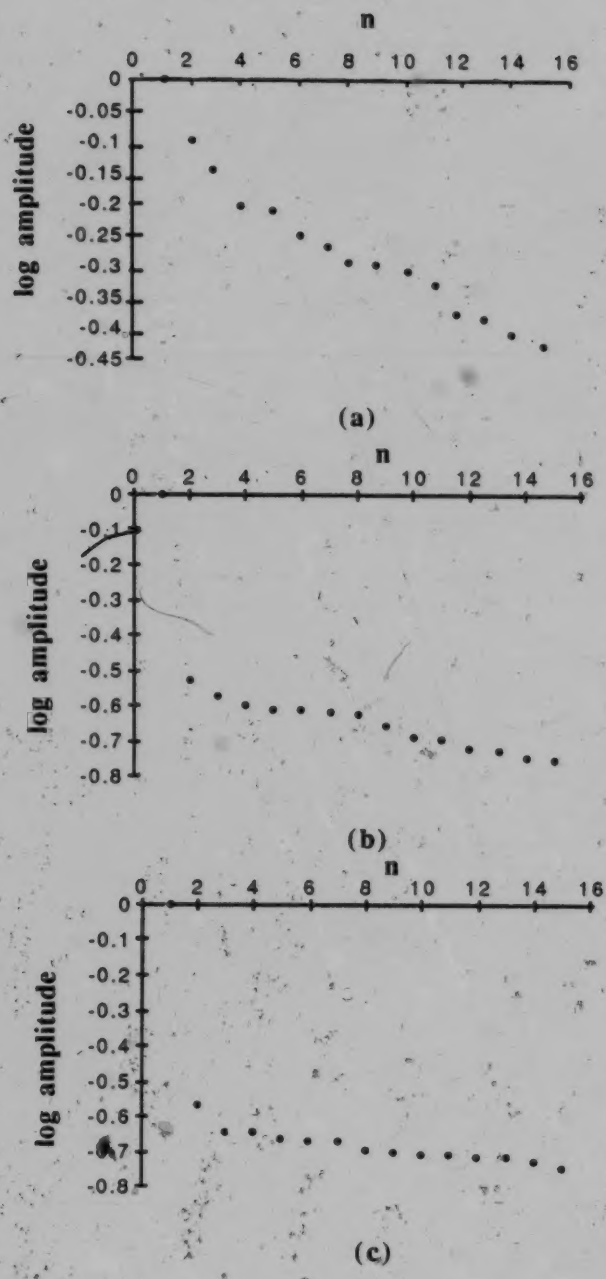


Figure 2.7 shows the results of the SSA. (a) is the spectrum for the catalog containing mainshocks of magnitude above 4. (b) is the spectrum for the catalog containing mainshocks of magnitude above 2. (c) is the spectrum for the catalog containing all the events of magnitude above 2.

Chapter 3: The Limit on Earthquake Prediction

Since the discovery of deterministic laws of physics, it was believed that predictions can be made in an unlimited way into future, at least in principle. This led to Laplacian determinism a century ago. But this is restricted at quantum scales by the Heisenberg uncertainty principle, a central dogma of quantum mechanics developed in this century. Recent development of chaos theory has now had a strong impact on large scale classical deterministic system; it shows that some dynamic systems are inherently unpredictable in long term. Although the future of a deterministic system is completely determined in principle by its past, any uncertainties, no matter how small, grow exponentially as time evolves. Prediction is, therefore, only possible in short term.

I believe that the dynamics of the lithosphere is a system of this kind. Although the analysis in the last chapter indicates that the dynamics of the lithosphere may have low dimensionality, and therefore should be treated as a deterministic system, it is not certain that the evolution of the lithosphere can be predicted far into future. The predictability of evolution of the lithosphere state encompasses all the complexities of understanding the dynamics of the lithosphere. The determination of the state of the lithosphere dynamics at a certain time is so difficult that it greatly diverts our attention from the fundamental dynamics. Taking advantage of recent development in non-linear dynamics, we can overcome apparent complexities in the dynamics of the lithosphere, and concentrate on the geometric properties of a low dimensional attractor for the dynamics. We therefore can possibly study the predictability of the lithosphere dynamics without having complete measurements and understanding of the dynamics of the lithosphere.

The discussion in the following sections are basically built on the works of Robert Shaw (1985) and others (eg. Farmer, 1982). Their view of predictability is the ability of a system to carry stored information into future. In applying such an idea to the dynamics of the lithosphere, I will again consider the earthquake catalog as an output of the dynamic system, and I will examine the Kolmogorov entropy of the system, and thereafter, determine the upper limit of time on earthquake prediction.

3.1 Information theory and dynamics

The information theory of communication was originated by C. E. Shannon in 1940's (eg. Shannon and Weaver, 1962). A communication system has three essential parts: transmitter, channel and receiver. Information is conveyed from the transmitter via the channel to the receiver. Statistical distributions are used in this theory to describe possible transmitted and received messages. The properties of the communication channel are described by a conditional distribution $P(y \mid x)$, which gives the probability of receiving message y given that message x was transmitted.

The idea of extending Shannon's communication system theory to a dynamic system is due to Shaw (1985). He views a dynamic system as a communication system which communicates some of the information about its past state into the future. The past and future variables in this case are the same set of coordinates at two different times. Figure 3.1a depicts the idea how the past of a dynamic system is related to the future. In a purely deterministic system the state is described as a point which represents an infinite amount of information, and is unphysical in this sense. A way around this difficulty is to apply a partition to the domain of the continuous variables. Figure 3.1b schematically shows the idea. The initial measurement finds that the state is contained in the little block s_i, and at a later time the state may fall over several blocks. The probability of finding the state at s_j is given by a conditional distribution $P(s_j \mid s_i)$, which shows the causal connection between past and future

If a state s_k is associated with the probability P_k, the information that we have gained by learning the system is in the state s_k, is defined as

$$I(s_k) = -\log P_k$$

The smaller the probability associated with the state s_k, the more information we will gain after learning the system is actually in the state s_k. Therefore, $I(s_k)$ measures the uncertainty associated with the state s_k. The binary base is generally used for the base of the logarithm. The unit of the amount of information is call a bit, where one bit is the

amount of information we can gain by answering one yes or no question. For a certain distribution $\{P_k\}$, the average amount of information is given by

$$H(S) = \overline{I(s_k)} = -\sum_k P_k \log P_k$$

where S is a random variable defined over the sample space. H(S) quantifies the amount of information gained when we learn that a variable is distributed as $\{P_k\}$, relative to a flat distribution. Analogous to $I(s_k)$, H(S) represents the expected value of the uncertainty associated with our probability scheme. It is also called entropy for, just as in thermodynamics, H(S) measures the level of our ignorance about the state of the system.

If the state variables at a past time impose any constraint on the future state of the system, then they must store some information about the system, and thus permit a certain degree of predictability. Consider a dynamic system which relates its future states s_j to its past s_i through a transition probability $P\{s_j \mid s_i\}$. For each past state s_i, there is a probability $P\{s_j \mid s_i\}$ of evolving to a particular state s_j through the dynamic development of the system. According to the discussion above, one may ask for the quantification of the amount of information stored in the past state variables that can be transmitted to the future. In another words, how many bits of information do we obtain by knowing that the future state s_j corresponds to the past state s_i, when we know the overall probability of s_j happening along with different s_i? A measure for the amount of information passing from past to future is given as

$$I(s_j \mid s_i) = \log \frac{P\{s_j \mid s_i\}}{P\{s_j\}} = \log \frac{P\{s_i \mid s_j\}}{P\{s_i\}} = \log \frac{P\{s_j, s_i\}}{P\{s_j\}P\{s_i\}}$$

Imagine two observers with access to the same system at two different times. The observer making the first observation could attempt to communicate to the later observer by putting the system in a particular state; the system then carries the information to the later observer. The later observer, however, knows nothing of the system state at the earlier time. His a priori knowledge that the system was at the state s_i earlier is the marginal probability $P\{s_i\}$. The a posteriori knowledge of the observer after knowing that the system is at the state s_j, is based on the conditional probability of being at the state s_i

earlier, that is, $P(s_i \mid s_j)$. Therefore, the gain of information is the logarithm of the ratio of $P(s_i \mid s_j)$ and $P(s_i)$.

The information gained after we learn that the past state s_i and the future state s_j are distributed as $P(s_i, s_j)$, is given as the average of the amount of information $I(s_i \mid s_j)$ we gained for the system passing from each particular past to future.

$$I(S_i \mid S_j) = \sum_{i,j} P\{s_i, s_j\} I\{s_i \mid s_j\}$$

$$= \sum_i P\{s_i\} \sum_j P\{s_j \mid s_i\} \log \frac{P\{s_j \mid s_i\}}{P\{s_j\}}$$

The subscript i indicates the past state and j indicates the future state. It will be zero when $P(s_j \mid s_i) = P(s_j)$, meaning the future and past are statistically independent.

3.2 Minimum information distribution and the earthquake cycle

Information is a relative concept, and is measured relative to our a priori knowledge of the system state. Lacking any measurement on the system, our a priori knowledge of the system state at a particular time is its asymptotic probability distribution. This special distribution represents a minimum knowledge of the system state. Shaw (1985) has brought this concept into the study of dynamic systems, and argued that the minimum information distribution is the base line against which all the information we have about the system should be measured. If our knowledge about the system differs from minimum information distribution, we are eligible to make a prediction to a certain degree.

A simple example to illustrate this concept is a harmonic oscillator with known energy but unknown phase (Figure 3.2). The probability of finding the oscillator at some position is proportional to the inverse of the velocity at that point:

$$P(x) \propto \frac{1}{\sqrt{1 - x^2}}$$

In the expression $I(S_i \mid S_j)$, the average information transmitted, the marginal distributions describe our a priori knowledge of the system states. Since before we make

any measurement on the past and future system states, our a priori knowledge of the system state is its minimum information distribution, the average information transmitted is

$$I(S_j \mid S_i) = \sum_i \bar{P}\{s_i\} \sum_j P\{s_j \mid s_i\} \log \frac{P\{s_j \mid s_i\}}{\bar{P}\{s_j\}}$$

where $\bar{P}$ is the probability distribution corresponding to minimum information. This formula quantifies the average increase in our ability to predict the future state of the system when we learn its past. $I(S_j \mid S_i)$ is defined as the average information stored in the system.

For the dynamics of the lithosphere, our a priori knowledge of the system is the distribution of certain measurements related to earthquake occurrence. The information in this case is measured relative to those distributions, which are generally not uniform both in space and time.

The seismic cycle is often mentioned in describing different characteristics of seismicity in the periods before a strong earthquake, after a strong earthquake, and in the period between consecutive strong earthquakes. It refers to the characteristic time intervals between the strongest earthquakes in a given seismic region. However, it is not a cycle in the usual sense of the word since the periods between the strongest earthquakes are not equal and can deviate considerably from the average characteristic period. Such concept can also be extended to the spatial variation of the epicenters of the strong earthquakes. Seismic gaps usually display a repetitive pattern (Shimazaki and Nakata, 1980; Thatcher, 1984; Mogi, 1985).

The existence of the earthquake occurrence cycle in time and space has led to a number of attempts to forecast the occurrence time and location of future events. Data on large earthquakes have indicated that the recurrence time of earthquakes in the Nankai Trough in Japan and southern Chile is 100-200 years (Ando, 1975; Kelleher, 1972). Sykes and Quittmeyer (1981) show that the recurrence time can vary from as short as 35 years in some areas to more than 150 years in others. Some surveys on active faults show roughly constant intervals of repetitive occurrence of large earthquakes over a long period (Sieh, 1978a, 1978b, 1981; Wallace, 1977). The spatial regularities of earthquake

occurrence appear as gaps in the spatial distribution of rupture zones of large earthquakes in a seismic belt. Mogi (1979) has pointed out a certain regularity of occurrence of great earthquakes in space along southwest Kurile Trench-northern Japan Trench. The focal regions almost completely cover the seismic zone of this area without overlapping, and display a periodic pattern in space. The development of seismicity in space shows a remarkable pattern along the southern Kurile trench. A similar spatial pattern was also found in the major plate boundaries of the Pacific and the Caribbean (Kelleher, et al., 1973).

The phenomenon of a seismic cycle indicates that the distribution of earthquakes is not uniform either in time or in space. If the system is ergodic, the distribution will be independent of time, and therefore it describes the time independent part of the lithosphere dynamic system. The knowledge about the time and space distribution of earthquake occurrence we have learned is our a priori knowledge of the system. This is the most we know about earthquake risk for a certain region without intervening with any further measurements on the system, and it is the information we generally use to make a earthquake risk estimation. However, if we make any further measurement, we will acquire an additional information about the lithosphere state that we can use to make a prediction of the lithospheric state in future. Earthquake prediction therefore results from information gained by further measurements on the lithosphere. How far we can predict into future, however, will depend how fast the information will be lost as time progresses.

3.3 Rate of loss of stored information

One of the properties of a chaotic system is that the uncertainties or errors in measurements will increase exponentially. If the future state variables are calculated based on the system equations with present measured state variables as the initial condition, the information stored in the future state variables will reduce due to the increase of uncertainty. In another words, such a system is incapable of carrying all the information into future. Eventually all the information obtained in the initial measurement will be lost and the distribution of the state variables will relax to the asymptotic minimum information

distribution.

However, if we continue make new measurements to monitor the system, our knowledge about the system will be updated by information that would have been lost if the measurements had not been made. The information acquisition rate of an observer who makes a series measurements at time $t_1, t_2, \ldots t_n, \ldots$ should be same as the rate of information loss of the system. The Kolmogorov entropy provides a measure of this information acquisition rate.

Suppose that the phase space of the dynamic system is partitioned into small boxes, and the state of the system at time t_i is denoted by s_i, which corresponds to a small box in phase space. A history of n measurements of the system at time $t_1, t_2, \ldots t_n$ are then expressed by a string of symbols $(s_1, s_2, \ldots s_n)$, abbreviated s^n. The probability

$$P(s_1, s_2, \ldots\ldots s_n) = P(s^n)$$

describes the joint probability that the system state is at state s_1 when $t = t_1$, at s_2 when $t = t_2, \ldots$, and at s_n when $t = t_n$. The information needed to locate the system on a special trajectory $s_1, s_2, \ldots s_n$ if our a priori knowledge is $P(s^n)$, is given as

$$H(S^n) = - \sum_{s_1 \ldots s_n} P(s^n) \log P(s^n)$$

where sum is taken over all possible patterns of symbols of length n. Therefore, the additional information needed to predict which state the system will be in if we know the system was previously in the state s^n can be written as

$$\begin{aligned} \Delta H_n &= H(S^{n+1}) - H(S^n) \\ &= - \sum_{s_1 \ldots s_{n+1}} P(s^n) P(s^n | s_{n+1}) \log P(s^n | s_{n+1}) \end{aligned}$$

This means, that ΔH_n measures our loss of information about the system from time t_n to time t_{n+1}. ΔH_n should be independent of n if s_i is a complete measure of the system state at time t_i, i.e. there will be no increase of information about the system state upon learning the history of the system $s_{i-1}, s_{i-2}, \cdots$. If s_i is not the complete measure of the system,

which is caused by the imperfect projection of the system state to the measured variables, more information can be gained by considering the past history of the measurements. ΔH_n should be independent of n if n is made large enough. The average rate of loss of information is given as

$$K = \lim_{n\to\infty} \frac{1}{n\Delta t} \sum_{n=1}^{n} \{H(S^{n+1}) - H(S^n)\}$$
$$= \lim_{n\to\infty} \frac{1}{n\Delta t} H(S^n)$$

The quantity K is defined as Kolmogorov entropy and is due to Kolmogorov and Sinai (Kolmogorov, 1959; Sinai, 1959). It characterizes the degree of dynamical disorder of the system, or how sensitively the system depends on its initial state. Figure 3.3 schematically illustrates the motion of a dynamic system in three situations correspond to $K = 0$, $K > 0$ and $K = \infty$ respectively. As is well known, $K = 0$ is in an ordered system, K is infinite in a random system, and $K > 0$ is a chaotic (deterministic) system.

We know from the discussion in the last section that $I(S_j \mid S_i)$ measures the degree to which an observer at t_i can predict the state of the system at t_j. Suppose we start at time t_1, and denote $I(S_2 \mid S_1)$ by I_1. Such a sequence can go on with $I(S_3 \mid S_1)$ denoted by I_2, and so on. We then obtain a sequence of numbers I_1, I_2, ..., describing the ability to predict further and further into future (Shaw, 1985). At beginning when $j = i = 1$, the conditional probability $P\{s_i \mid s_j\}$ must satisfy $P\{s_i \mid s_i\} = 1$, therefore we have

$$I(S_i \mid S_i) = H(S_i)$$

This is the initial information we have about the system after we determine the system is at a certain state. Based on this information we are trying to predict the future states of the system. If the system is not chaotic, the next state will be uniquely determined by the previous one. In this case $P\{s_j \mid s_i\} = 1$, and we have $I(S_j \mid S_i) = H(S_i)$, the system will carry the initial information far into the future. If the system is chaotic, the system will tend to lose the information. As time tends to infinity, the initial state s_1 of the system will be dynamically uncorrelated to the future state s_n for an ergodic system, and

$$\lim_{n\to\infty} P(s_n \mid s_1) = P(s_n)$$

This implies that $I(\infty) = 0$, that is all the information stored in the system will be lost eventually. The general form of I(t) is illustrated in Figure 3.4. The curve is concave and approaches zero asymptotically (Shaw, 1985). Initially I(t) decreases at a linear rate given by the Kolmogorov entropy (Farmer, 1982).

Since the information decays linearly to begin with, the initial change of the stored information can be written as

$$I(t) = I(0) - K \cdot t$$

One should recall the information dimension $D_I = \lim_{n\to\infty} I(\varepsilon)/\log(1/\varepsilon)$ (chapter 2), where ε is the linear dimension of the partition and $1/\varepsilon$ therefore represents the signal to noise ratio. The above expression can be written as

$$I(t) = D_I \log\left(\frac{1}{\varepsilon}\right) - K \cdot t$$

The characteristic time is therefore given as (Farmer, 1982)

$$\tau = \frac{D_I}{K} \log\left(\frac{1}{\varepsilon}\right)$$

This characterizes the upper limit of the time over which the state of a chaotic system can be predicted. The initial data become useless after the characteristic time τ, and τ is only influenced logarithmically by the precision ε within which the initial state is located.

3.4 Estimate the upper limit of time on earthquake prediction

According to above discussion, the estimation of the upper limit of time for predictability of an earthquake becomes the problem of estimating the Kolmogorov entropy of the dynamics of the lithosphere. There have been numerical attempts to determine K based directly on the definition. Grassberger and Procaccia (1983b) have suggested a way

of estimating K by using a generalized correlation integral, which we will apply here to earthquake prediction. This approach has a high potential in implementation to a measured time series.

Consider the entropy series K_1, K_2, ... defined by

$$K_f = - \lim_{n\to\infty} \frac{1}{n\Delta t} \frac{1}{f-1} \log \sum_{s_1 \ldots s_n} P^f(s^n)$$

It is straight forward to see that K_1 corresponds to the Kolmogorov entropy. Generally we have following relation:

$$K_{f'} \leq K_f \qquad \text{for} \quad f' < f$$

Therefore, K_2 serves as an upper bound of the Kolmogorov entropy. It can be easily determined numerically.

Recall that the correlation integral is given as

$$C(\varepsilon) = \sum_i P_i^2$$

where P_i is the probability that the trajectory of the phase portrait will visit the i-th box. The summation over i runs through all the boxes which could be possibly visited in phase space. In the case of a history of states s_1, s_2, ... s_n, the correlation between the state strings can be similarly considered, and the correlation integral can be generalized to

$$C_n(\varepsilon) = \sum_{s_1 \ldots s_n} P^2(s^n)$$

$$= \lim_{N\to\infty} \frac{1}{N^2} \left\{ \text{number of pairs ij with } \sqrt{\sum_{m=1}^{n} (s_{i+m} - s_{j+m})^2} < \varepsilon \right\}$$

As n increases, the generalized correlation $C_n(\varepsilon)$ will decrease exponentially as

$$C_n(\varepsilon) \sim e^{-n\Delta t K_2}$$

K_2 therefore can be written as

$$K_2 = \frac{1}{\Delta t} \lim_{n \to \infty} \log \left[\frac{C_n(\varepsilon)}{C_{n+1}(\varepsilon)} \right]$$

In practice we do not need to follow the evolution of all the degrees of freedom. Since the phase portrait can be reconstructed from a single time sequence while still preserving geometrical invariants of the dynamics, $C_n(\varepsilon)$ can be calculated from a time sequence $\{x_i\}$ by

$$C_n(\varepsilon) = \lim_{N \to \infty} \frac{1}{N^2} \left\{ \text{number of pairs ij with } \sqrt{\sum_{m=1}^{n} |x_{i+m} - x_{j+m}|^2} < \varepsilon \right\}$$

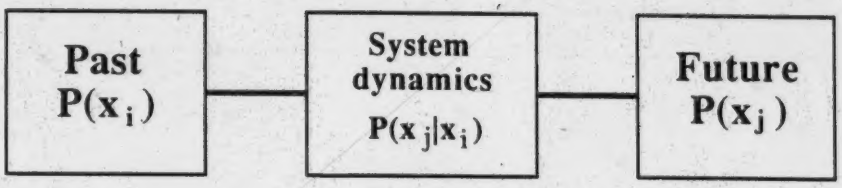
The analysis was done for the west coast of Canada. The average time interval between successive events is used for Δt in the calculation of the entropy K_2. The results shown in Figure 3.5 are the analysis for the earthquake catalog that contains the events of magnitude above 3, and has 1700 events. Figure 3.5a is $\log C_n(\varepsilon)/\Delta t$ against n. Different curves are the results calculated for different ε. $\log C_n(\varepsilon)/\Delta t$ is a measure of information, and as expected the information increase as we consider more historical events. The slopes of the curves, $K_2(\varepsilon,n)$, is the rate of information acquisition when we make more measurements, or the rate of information loss of the system. Figure 3.5b shows $K_2(\varepsilon,n)$ decrease as n increase, and finally tend to a constant value of K_2 entropy 0.6 bit/year. Different ε does not effect $K_2(\varepsilon,n)$ converges to the common value as expected. Figure 3.6 shows the same analysis for the catalog that contains the events of magnitude above 2, and has 3160 events. It shows that K_2 is 1 bit/year. A higher entropy in the latter case is probably due to the reason that the catalog becomes noisier when more small events are included.

The value of D_2 calculated in the last chapter can be used as the estimate of the information dimension. The earthquake catalog can be polluted with various kinds of noise as discussed in the last chapter. We do not have a clear guideline on how to choose the value of the signal to noise ratio for the catalog of the west coast of Canada. I suggest the value of 10 as the signal to noise ratio. Such choice is not significant to the determination of the characteristic time because it depends on the signal to noise ratio logarithmically. By using 0.6 bit/year for the entropy K_2, the characteristic time, which is the upper limit of

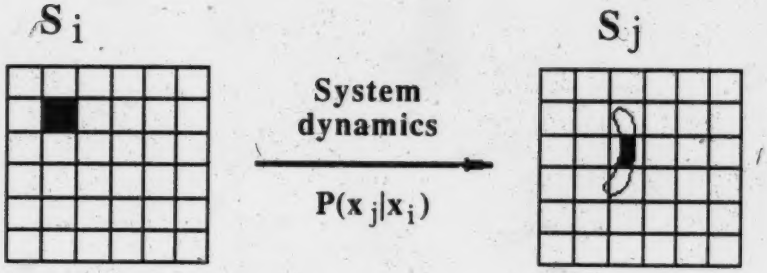
time for the earthquake prediction is estimated as 20 years for the Vancouver Island area. This upper limit indicates how far into the future we can predict an event of magnitude above 3 in an ideal case, in practice however it could be much less.

The catalog for the west coast of Canada presents with marginal homogeneity, and contains minimum required number of events for the statistical analysis discussed above. Therefore, I cannot conclude from this analysis whether the results obtained are stable to the drawbacks of the catalog. Such analysis can be done for the catalog contains more seismic events and have a longer time coverage.

I should emphasize here that according to the conceptual frame work discussed in the last chapter, the attractor of the lithosphere dynamics can be reconstructed from an earthquake catalog, therefore, it is important to notice that the upper limit of time on earthquake prediction obtained from the analysis of seismicity data for the west coast of Canada is not only a limit on the earthquake prediction by using the earthquake catalog, but is a fundamental characteristic time of the system. This characteristic time is intrinsic to the attractor of the lithosphere dynamics, and characterizes the predictability of the system. It can not be improved by better understanding of the lithosphere, or making more measurements on other geodynamical observables. All attempts at earthquake prediction are subject to this limit.



(a)



(b)

Figure 3.1 (a) illustrates the idea that the past of a dynamic system relates to the future in the same as a communication channel. (b) illustrates the idea of applying a partition to the domain of continuous variables so as to the deterministic dynamics can now map an image of each little block (Adapted from Shaw, 1985).

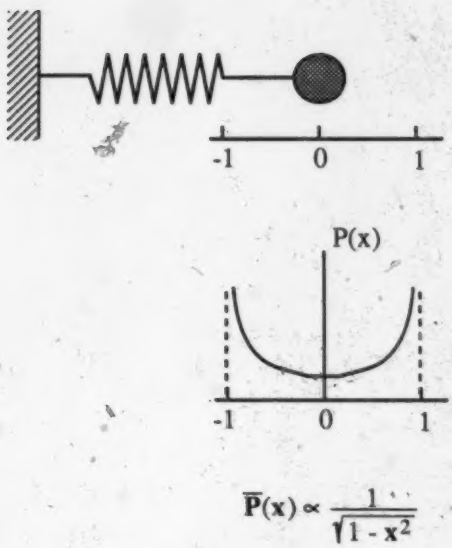


Figure 3.2 is an example to illustrate the concept of minimum information distribution, which is the probability of finding the position of an oscillator without knowing any information about its phase (Adapted from Shaw, 1985).

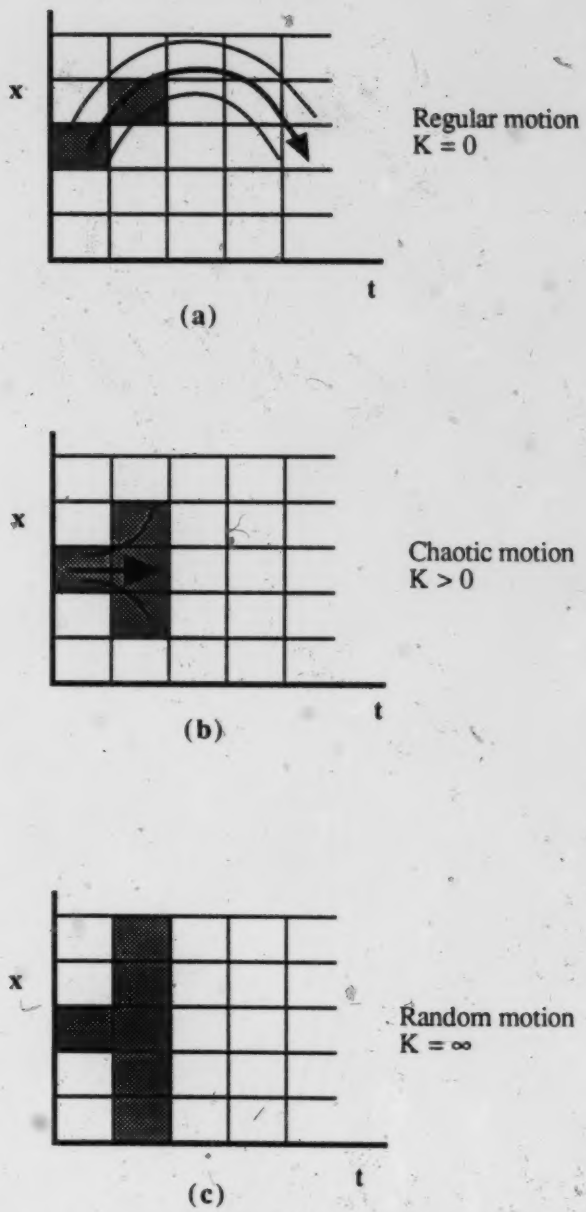


Figure 3.3 schematically illustrates the motion of a dynamic system in the extended phase space. (a) is regular motion; the next state is uniquely determined. (b) is chaotic motion; the trajectory spreads. (c) is random motion; it can be in any state in the next time. (Adapted from Schuster, 1984)

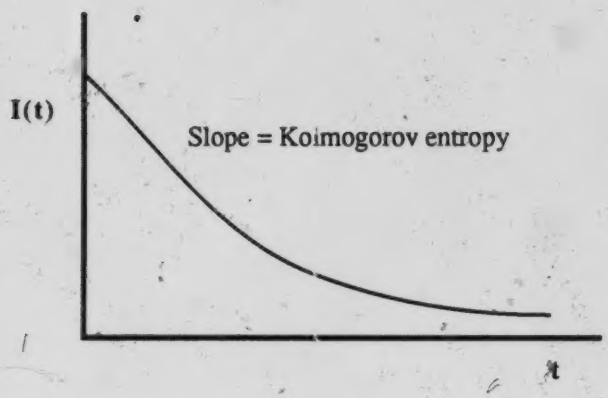


Figure 3.4 illustrates the typical behavior of I(t) for a chaotic attractor. Initially I(t) decreases at a linear rate which equals Kolmogorov entropy (Adapted from Farmer, 1982).

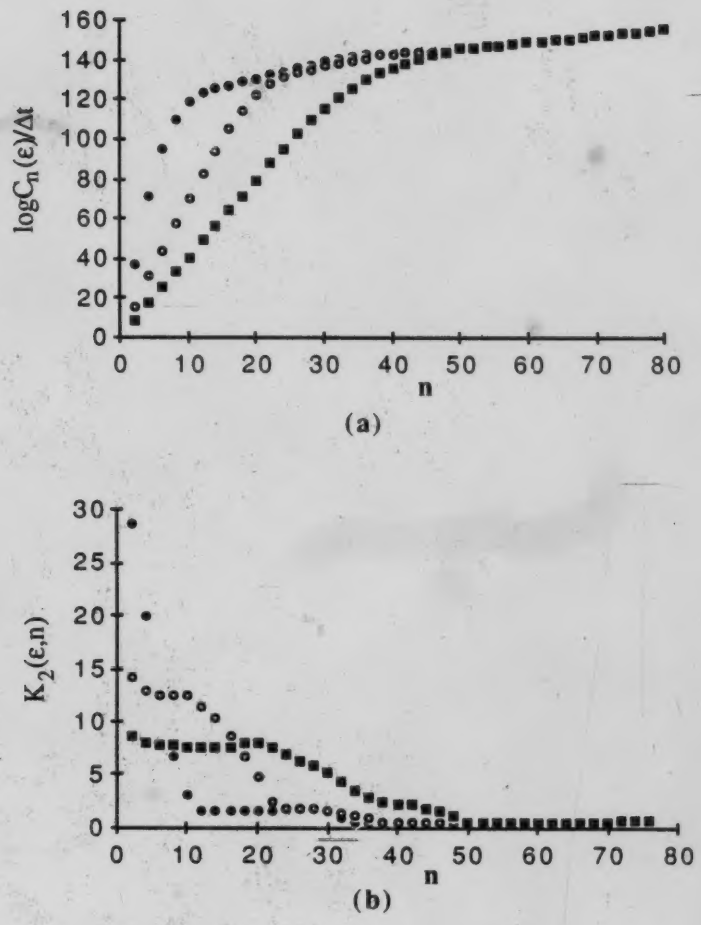


Figure 3.5 (a) shows $\log C_n(\varepsilon)/\Delta t$ versus n calculated for the catalog containing all events of magnitude above 3. The slopes of the curves $K_2(\varepsilon,n)$ are shown in (b), which describe the rate of information loss of the system.

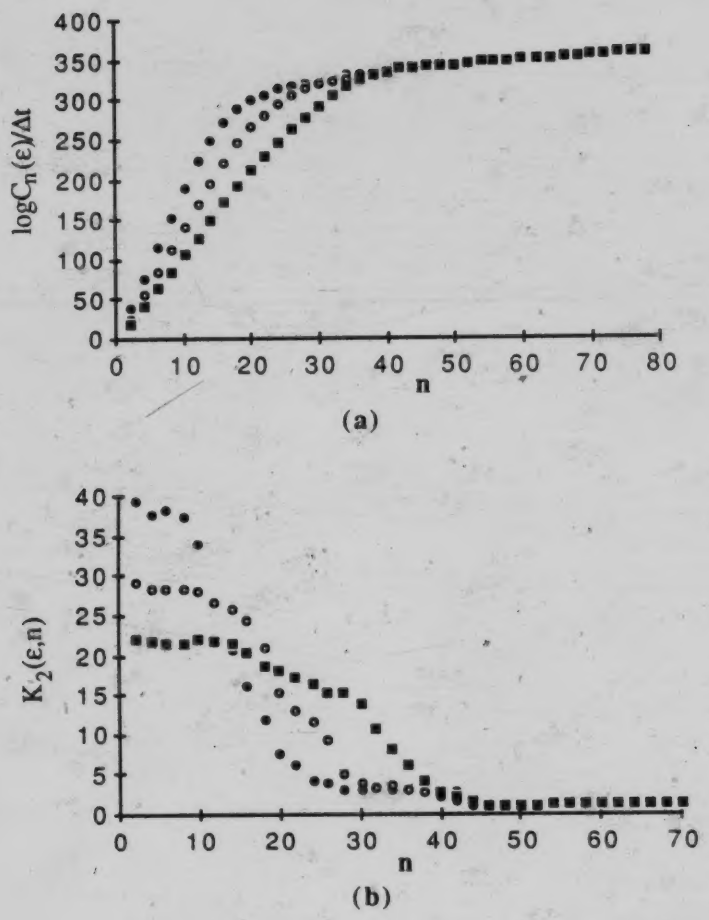


Figure 3.6 (a) shows $\log C_n(\varepsilon)/\Delta t$ versus n calculated for the catalog containing all events of magnitude above 2. The slopes of the curves $K_2(\varepsilon,n)$ are shown in (b).

Chapter 4: Detecting premonitory seismicity patterns

The analyses in the previous chapters suggested that the dynamics of the lithosphere are chaotic rather than random. Therefore we expect that there may exist an underlying order associated with a low dimensional attractor. The existence of precursors preceding strong earthquakes and world wide similarity of seismicity supports the argument that the dynamics of the lithosphere is not as random as it seems to be. This gives us a hope to make a prediction for future earthquakes. Predicting future earthquakes is a process of extracting the information about the present dynamic state of the lithosphere, which will impose certain constraints on the future development of the lithosphere dynamics. If the dynamics of the lithosphere is chaotic, the information associated with occurrence of future earthquakes can be extracted from the earthquake catalogs. The problem of earthquake prediction then become a matter of finding a way of decoding earthquake catalogs.

Although my analysis suggests that a low dimensional attractor exists in lithosphere dynamics, its geometrical properties are by no means well understood. Generally, we need to have complete knowledge about the attractor in order to understand the dynamics of the lithosphere, and thereafter to understand occurrence of earthquakes. However, things become much easier in the case of earthquake prediction. We only need to determine whether the lithosphere is in a certain dynamic state which is believed to precede a strong earthquake, and to discriminate these states from others. Since the dynamic properties of the system will be reflected in geometrical structure of the attractor, those lithosphere states preceding a strong earthquake may correspond to a region in the attractor characterized with a special geometric pattern. Pattern recognition techniques may clarify the complexity of identification of such regions and it is possible to automate this function using computers.

Two pattern recognition algorithms for long-term earthquake prediction have been developed by Soviet scientists in the last decade. The precursors preceding strong earthquakes are sufficiently well defined in these algorithms to be tested statistically. World wide application of these algorithms continues to add credibility to the technique.

These pattern recognition algorithms are basically empirical methods motivated by some abnormal seismic activity found in earthquake catalogs. Linking these patterns to the

geometric patterns on the attractor of the lithosphere dynamic system is a small conceptual leap. The advantage of this consideration is to introduce the theory of the dynamics of the lithosphere into the algorithm. We therefore can take advantage of recent achievements in non-linear dynamic theory in our further study of the algorithm.

4.1 General Idea of Pattern Recognition

Many important applications of pattern recognition can be characterized as either waveform clarification or clarification of geometric figures. In either case the subjects can be described as a vector $\mathbf{x}=(x_1.....x_n)$. Thus, the purpose of the pattern recognition becomes to discriminate between two sets of points in an n-dimensional space. For a simple two-dimensional case as shown in Figure 4.1, we can set up a boundary, $g(x_1, x_2) = 0$, between these two distributions if we know two distributions of vectors $\mathbf{x}$ from past experience (Fukunage, 1972). We then can decide which group the vector $\mathbf{x}=(x_1, x_2)$ belongs to by examining whether $g(x_1, x_2) > 0$, or $g(x_1, x_2) < 0$. The discriminant function $g(x_1, x_2)$ is a classifier which has to be determined according to our knowledge of the distribution of the vector $\mathbf{x}$.

With a large number of primitive measurements, high dimensionality makes human judgement impossible; human beings usually make a classification based on a small number of features of an object. The most common treatment for such a case is to map the primitive n-dimensional space to a lower dimensional space. The mapping, of course, should be carried out without severely reducing the class separability. Obviously, the selection of such variables is important and strongly affects classifier design, ie. if the variables show significant differences from one class to another, the classifier can be designed more easily with better performance. Therefore, the selection of these variables is a key problem in pattern recognition.

Unfortunately, in many applications of pattern recognition, such features are not linear functions of original measurements, but highly nonlinear functions. Because of the complexity in mathematics, we do not have a general theory to generate mapping functions systematically and to find the optimum one. However, if a set of data is associated with a

low dimensionality, the existence of a minimum number of effective features is then guaranteed. Due to the lack of systematic treatments in mathematics, effective features primarily have to be found in practical application by the designer's knowledge of the physical system. A simple way of testing whether the feature functions selected by the designer are informative in separating two classes of objects is to examine the one dimensional distribution histogram of the functions for different objects. The more different the distribution is for different objects, the more informative the function is in separating two class of objects.

In the problem of earthquake prediction, we are interested in choosing those features which are most effective for separating two classes: one is the objects N (standing for Not dangerous or Nyet) consisting of time periods within which there will be no earthquakes in a particular region, another is the objects, denoted by D (standing for Dangerous or Da), during which earthquakes will occur in that region (Keilis-Borok, et al., 1986). If we select a feature function P which varies in an interval $[P_0, P_n]$, we may compare distributions or the values of P for objects N and D in order to find out whether this function is useful for discrimination between N and D. Figure 4.2 shows an example of such a one dimensional distribution. Within the interval [2, 7], the distribution of the values of function P is quite different from one class to another. It implies the function P is effective in discriminating class N from D. If, on the other hand, the distribution does not show significant differences, then the function P is not useful.

The value of the function P in discriminating class N and D can be quantified by defining an informative function (Keilis-Borok, et al., 1986)

$$I_P = (1 - \varepsilon)\frac{n_D(\Delta)}{n_D} - \varepsilon \frac{n_N(\Delta)}{n_N}$$

where $n_D(\Delta)/n_D$ and $n_N(\Delta)/n_N$ are empirical cumulative distribution functions of P for objects D and N. n_D is the total number of points in class D, and $n_D(\Delta)$ is the number of the points in class D with $P \geq \Delta$. n_N and $n_N(\Delta)$ are same for class N. The parameter $0<\varepsilon<1$ determines the relative costs of failure to predict and false alarm. ε and Δ are parameters can be adjusted to maximize I_P.

The set of functions used in the following analysis is the standard set of functions used in the algorithm M8. Worldwide experiments indicate those functions are informative at distinguishing the Time of Increased Probability (TIP) for occurrence of a strong earthquake from the rest of the time domain. With such a set of functions, the pattern recognition technique can be performed. The idea of this approach is illustrated in Figure 4.3. The vertical dashed line shows a sliding moment of time. At each moment we look back in time and define several characteristics (traits) of the earthquake sequence within the sliding time windows, indicated by the horizontal lines. Each of these characteristics is represented by several functions, and these functions are considered as the components of a vector. The variation of the value of a vector will indicate the time of increased probability.

4.2 Definition of the Functions of Algorithm M8

Pattern recognition requires a separation between mainshocks and aftershocks. They are strongly related and the aftershocks are considered as dependent events of the mainshocks. Aftershocks are generally defined empirically with some selected time and space windows. If two events are denoted by i and j, the event i is an aftershock of the event j if and only if the distance between their epicenters is less than $\mathbf{R}(M_j)$, the time difference $t_i - t_j \leq \mathbf{T}(M_j)$, and, $M_i < M_j$. $\mathbf{T}(M_j)$ and $\mathbf{R}(M_j)$ are empirical functions which can be viewed as simple box car windows containing obvious "mainshock-aftershock" clustering (Lamoreaux 1982). Gardner and Knopoff (1974) suggested logarithmic relations between time **T** and magnitude M_j; and distance **R** and M_j. Inspection of the events in the Western Canada test data that we used clearly indicated aftershocks at larger distances than considered in the California area by Gardner and Knopoff. As a consequence we increased the maximum distance at which an event could be considered an aftershock by 50% (Table 4.1). After this operation, we can obtain a main shock catalog, with a certain number of aftershocks following each events.

The algorithm M8 involves 7 functions, which form a vector with seven components. In other words, the high dimensional space of primitive measurements is mapped to a 7-dimensional space. Those functions are divided into four groups, each of

which consists of functions defined by the same formula, but different in the values of numerical parameters. Definitions of the functions are as follows:

The **level of seismic activity** is the number of main shocks that occurred in an interval s before a time t and had a magnitude greater than a lower limit $\underline{M}$. We refer to this count as $N(t ; \underline{M},s)$. A more sophisticated measure of main shock activity calculates a weighted form of this sum where the weights are proportional to the size of the earthquake. Instead of simply summing the events we calculate $S = \sum_i 10^{\beta(M_i - \alpha)}$. For $\beta=b/3$ this is a measure of the rate at which the radius of circular fractures is generated and for $\beta=2b/3$ the rate at which fracture area is generated. Here b is the usual coefficient in the energy magnitude relation $\log(E)=A+bM$. We carry out this process for all events whose magnitude M_i is between $\underline{M}$ and an upper limit $\overline{M}$ ($\overline{M} = M_0 - 0.1$) and call the result $S(t; \underline{M}, \overline{M}, s, \alpha, \beta)$. (The parameter α is often inserted to normalise the functions.)

The **spatial concentration** is the combination of the pair of functions described above to produce a measure of the way seismicity concentrates in space.

$$Z(t; \underline{M}, \overline{M}, s, \alpha, \beta) = \frac{S(t; \underline{M}, \overline{M}, s, \alpha, \beta)}{(N(t; \underline{M} s) - N(t; \overline{M} s))^{2/3}}$$

defines the ratio of the average radius of a fracture (S/N) to the average spatial event separation for all events between $\underline{M}$ and $\overline{M}$, if they are uniformly distributed in the region $(N^{1/3})$.

The **long-term variation of seismicity** is defined for the situation where the rate at which main shocks are produced is not uniform. If we assume that on the average the rate is constant, the function

$$L(t; \underline{M}, s) = N(t; \underline{M}, t-t_0) - N(t; \underline{M}, t-t_0-s) \frac{t - t_0}{t - t_0 - s}$$

measures the degree to which the generation of main shocks deviates from a long term linear trend.

These three functions depend explicitly on the parameter $\underline{M}$. If we expect to compare results with those in other areas we must find some way to normalise the result

and we do that by adjusting $\underline{M}$ so that the average yearly rate of occurrence of main shocks is in one case ten per year for one set of functions we designate as N_1, L_1, and Z_1 and is as close to twenty per year as possible for another set N_2, L_2, and Z_2. Thus, we considered six functions in three groups.

The **clustering in time and space** is the count of aftershocks that occur within a few days of a main shock. The occurrence of abnormally large clusters of aftershocks of main shocks has been proposed as a precursor of strong shocks.

$$b(t; \underline{M}, \overline{M}, s, M_a, e) = \max_{(i)} b_i(M_a, e)$$

Here $b_i(M_a, e)$ is the number of aftershocks of $M \geq M_a$ for the i-th main shock at the time interval $(t_i, t_i + e)$; maximum is taken on the main shocks with $\underline{M} \leq M_i \leq \overline{M}$ in the time interval (t-s, t).

We can now consider the earthquake stream to be characterized at a time t by a vector of functions deduced from the preceding time period s and a selected square region. Our problem is to deduce, by pattern recognition, whether such a time belongs to a TIP, the times of increased probability. The duration τ of such a TIP, the time within which a strong earthquake is expected, is taken to be the 5 years used in all recent applications of M8 world wide.

In order to declare a TIP at time t we require that over a period ϕ, prior to and including t, three of the groups $\{N_1,N_2\}$, $\{L_1,L_2\}$, $\{Z_1,Z_2\}$, and $\{B\}$ contain functions that have extreme values and that at least four of the functions N_1, N_2, L_1, L_2, Z_1, Z_2, and B have extreme values. In this sense extreme values are values in the upper 10% of the range achieved by the function from the start of the catalog to the time of occurrence of the first large ($M>M_0$, the prediction threshold) event. Once this condition occurs the interval $[t,t+\tau]$ is declared to be a TIP.

4.3 Diagnosing the premonitory seismicity patterns near Vancouver Island, Canada

The algorithm M8 was tested for the Vancouver Island region. The following

analysis was directed by V. I. Keilis-Borok, from O. Yu Schmidt Institute of Physics for the Earth, Moscow, while he visited the University of Alberta at summer of 1987. D. Brown, E. Nyland and D. H. Weichert also participated the work. The results of the analysis is published in the paper "Premonitory seismicity patterns near Vancouver Island, Canada" in *Tectonophysics*, Vol. 167, 1989. What follows is an extract from this paper.

An area such as the southwest coast of Canada can be considered as a collection of overlapping regions defined by their seismicity. We do such regionalization by first dividing the area into rectangular cells defined by equal intervals of latitude and longitude. The square region centered on this event is then defined to have a side of length $R = e^{M_0 - 5.6}$ in degrees of latitude. Each such proposed region is then examined to see whether over the time span covered by the catalog an average of at least 10 main shocks occur per year. Those proposed seismic regions that do not contain 10 main shocks per year are discarded and the M8 algorithm is applied to the rest to determine whether a TIP exists in any of them.

The seismicity of the west coast of Canada has been described by many authors (e.g. Milne, et al., 1978; Basham, et al., 1982; Basham, et al., 1985). In the area we studied, from 118°W to 130°W and from 45°N to 51°N, we merged the seismicity catalog of the Geological Survey of Canada with the catalog from the National Earthquake Information Service in the US. Particularly for the earlier years the Canadian catalog contained many significant earthquakes as far south as 45°N. The duplicate events were removed manually. Usually NEIS data was accepted in the United States and GSC data was accepted in Canada. If only one agency reported a magnitude the report of that agency was accepted. Events for which no magnitude was reported were assigned zero magnitude, and if a preferred magnitude was given it was used. For these data that is usually M_L. We also removed those events since May 1980 that were clearly related to the Mt. St. Helen's Volcanic Eruption. The data is summarized in table 2, a time magnitude histogram, and in figure 4.4, a map showing the largest magnitude earthquakes that occurred in regions 1 degree by 1 degree.

Patterns in space and time can be distorted by changes in the observing networks and associated changes in the degree of completeness of the seismicity catalogs at lower

magnitudes. Evolution of the station configuration in the Vancouver Island area, and the associated increase in completeness of the catalog at small magnitudes has been discussed by Basham et al., (1982). The pattern recognition approach used in this work does not require as stringent a completeness condition as discussed there. Thus table 4.2 indicates that since 1952 a sufficient number, but certainly not all of the magnitude 2.5 earthquakes are present in the catalog for some parts of the test area. The effect of missing small magnitude events seems to predominate in the function that counts aftershocks and probably reduced its effectiveness as a diagnostic in Vancouver Island.

Similar considerations apply to the time variable location accuracy of the catalog. The spatial resolution of the analysis in this paper is about 0.5°, reflecting epicentral accuracy. Focal depths are the least certain, or nonexisting in some subareas, however the only application of focal depths is an exercise to determine whether patterns in shallow seismicity precede an intermediate earthquake.

In the analysis of the subcatalog discussed above we used primarily the standard parameter set developed for M8 by an analysis of a large number of regions world wide. We adjusted the threshold for prediction M_0, and the start time of the catalog t_0, to satisfy completeness considerations (Table 4.2) as required to normalise the Vancouver Island earthquake flow to others used in M8 and applied local judgment to adjust the windows used to eliminate aftershocks.

Table 4.3 shows the 9 strongest earthquakes of the area and indicates that the threshold for prediction should be M_0=6.5. This choice is by no means unique but is a stable choice; small variations do not seriously affect our conclusions. In order to select earthquakes for prediction we chose, before beginning the prediction process, all main shocks of magnitude greater than 6 that occurred since 1957. Since a main shock of magnitude 6.4 that occurred in 1971 near 50°N and 128°W was followed within 1 year by one of magnitude 6.0 near 49°N and 129°W we noted the both events as candidates for prediction. The 6.0 magnitude is substantially lower than our threshold for prediction so we place no great emphasis on the results for this event. We excluded two large aftershocks of magnitude 6,4 and 6.1 that occurred in 1972 and 1978, two large events of magnitude 6.3 that occurred in 1954 and 1956 are not candidates for prediction since the

preceding seismicity catalog is inadequate. Two events of magnitude 6.0 that occurred in 1958 and 1961 were ignored on the grounds that their magnitudes were both small and their time of occurrence was uncomfortably close to 1957, the beginning of a useful prediction data set. We therefore set as our goal the prediction of one event near Seattle and three west of the northern end of Vancouver Island.

Figure 4.5 shows the TIPs that resulted from the analysis of all the seismicity data available to us. The heavy lines under the time scale indicate TIPs, the light lines point to the center of the region in which the TIP occurred. The three events west of the northern end of Vancouver Island, of magnitude 6.0, 6.4, and 6.7 are associated with three regions in which TIPs occurred. Although the prediction for the magnitudes 6.0 and 6.4 should strictly be considered false alarms, we accept them as being within the uncertainty of magnitude determinations. The magnitude 6.7 was preceded by a TIP in two regions and preceded by a 5 year long TIP, which expired two years before the event, in another region. It was not seen in the data from a region due east of the event. The magnitude 6.0 event just to the north west was seen in three TIPs and not seen on the region to the east. The magnitude 6.4 event resulted in a TIP for both regions that contain it.

The magnitude 6.5 event near Seattle occurred at a depth of 59 km but the activation that preceded the event was largely shallow. Nevertheless the existence of four successful TIPs, one almost successful TIP, and only one region in which a TIP did not appear argues strongly for coupling between the shallow and deep seismicity in this area.

The quality of prediction is estimated by space-time volume of all TIPs in percentage of total space-time volume considered. The space-time volume is area (km^2) multiplied by time (years). The analysis shown in figure 4.5 gives a quality estimation of 30%.

The regions considered here are fairly large (figure 4.6) and it is an interesting speculation whether their overlap or their sum defines the region of high risk. We note that the events associated with the TIPs all occurred in that part of the activated regions common to each TIP. The part of the time domains common to all TIP is shown in the bottom of figure 4.6 and contained both the 6.5 and the 6.7 event. This does not prove that the intersections of the separate domains will in fact refine the predictions, but it is certainly

suggestive.

In order to explore more closely the potential of shallow tectonic activity to predict deeper events we attempted to separate the catalog into essentially crustal earthquakes and events related to subduction. Depth determinations in this area are not adequate to a careful separation so we chose to discriminate by considering only events of depth known to be less than 25km. This is clearly a set of lithosphere events but does not include all of them. The results of an analysis of this subset are shown in figure 4.7; there are not enough events in the remainder to allow application of M8. As expected, since most seismicity near Seattle is shallow, the TIPs around the Seattle event are unchanged. One region to the south of the 6.7 event is lost due to insufficient events in its subcatalog but the others remain viable TIPs.

The analysis that exploits the entire data set is associated with two TIPs for which large events have not yet occurred. In one case the seismicity data indicate that a Time of Increased Probability of an earthquake of magnitude greater than 6.5 began in the middle of 1982 and will expire in 1991. If past experience is a guide this TIP is associated with the northern end of Vancouver Island. The other TIP is associated with the northwestern corner of the United States. It began in the middle of 1983 and also expires in 1991.

Some traits of the catalog contribute to the TIPs and others do not. In figure 4.8 we illustrate the time intervals that created TIPs for 5 windows. The stars indicate the occurrence of extreme values for functions on the left at times indicated at the bottom of the plots. The shaded regions are the time intervals in which the catalog activation was sufficient to establish a TIP. Table 4.4 is a different way to summarise these results.

As indicated in figure 4.5, the M8 algorithm is not particularly effective in specifying the location of the anticipated earthquake. We and others have found it helpful to examine activation and quiescence on a more detailed scale and in figure 4.9 we show some of the regions subdivided in 256 cells. We have examined each cell to determine if significant activity occurred in it during the 6 years prior to the large event. If no such activity occurred the cell was identified as quiescent. A collection of quiescent cells bounded by activated cells may indicate the area in which the next strong event will occur. In each of the four cases, the events plot on the boundary between the quiescent and

activated areas, but the regions in the northwest show a better defined area of quiescence than those in the southeast.

The same process can be used to refine the time of occurrence of the anticipated earthquake. The quiescence in time will help us to refine the time when an earthquake will occur. Table 4.5 is a histogram of event number against magnitude and time for the magnitude 6.7 and 6.5 events. A quiescent period occurs before the magnitude 6.7 event and a courageous analyst would have predicted the event when the activity began to increase. The histogram of magnitude 6.5 event does not show such an effect as clearly.

4.3 Discussion

It is scarcely surprising that the seismicity of Vancouver Island bears a strong similarity to twenty areas previously tested by Keilis-Borok and his colleagues (Gabrielov, et al., in prep.; Keilis-Borok, 1990). Apparently identifiable patterns exists in the seismicity of Vancouver Island prior to at least four large seismic events. These premonitory patterns alone are not convincing, but the agreement of this phenomenon with many cases world wide suggests that existing an underlying order in the dynamics of the lithosphere, which governing earthquake occurrence. Since these patterns are universal, not depend on local geological structure and tectonic regime, it implies that they are intrinsic to the dynamics of the lithosphere.

Despite the success in world wide application of the algorithm, the physics behind those patterns still has not been understood. We have reason to believe that these patterns are linked to geometric properties of the attractor that exists in the lithosphere dynamics, but how they are related to each other qualitatively or quantitatively is still not clear. Nevertheless, linking the two together is not only conceptually inspiring in assuring us that earthquake prediction can be done as long as we dig deep into earthquake catalogs, but also constructive in suggesting us in which direction we could make a further improvement on the algorithm, which can not be provided by the algorithm itself. According to the discussion in previous chapters, we can make following suggestion for further improvement on the algorithm:

1) The dimension of the attractor is an important parameter which characterizes the dynamics. It can vary from region to region depends on the dynamic state of the area. If the dimension is n for a study area, we should choose at least n independent feature functions in order to obtain the maximum information about the dynamical state of the lithosphere. If the seismicity patterns are associated with some geometric properties of the attractor, we should choose at least 2n+1 independent feature functions in order to recover those properties of the attractor. Using too many feature functions will not retrieve more information but only increases computing time.

2) The feature functions are defined according to a certain time window. The rate of lose of information indicates the existence of a characteristic time, which gives a upper limit for selecting the time window, while the dimension of the attractor gives a lower limit. If the time evolution of these feature functions is examined in the algorithm, the sampling rate of the feature functions should be less than the characteristic time in order to resolve the determinism.

3) Most feature functions selected in the algorithm are the projections of a mainshock catalog. However, the analysis in chapter 2 indicates that aftershocks contain significant amount of information about the dynamics of the lithosphere. They deserve more consideration in the algorithm.

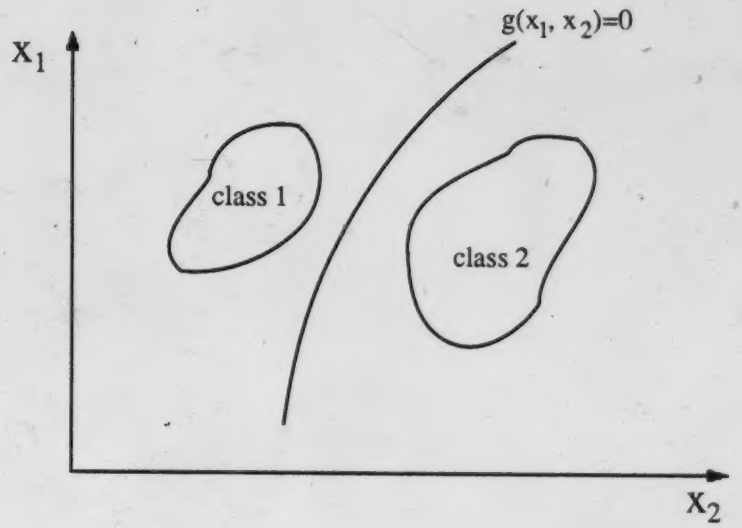
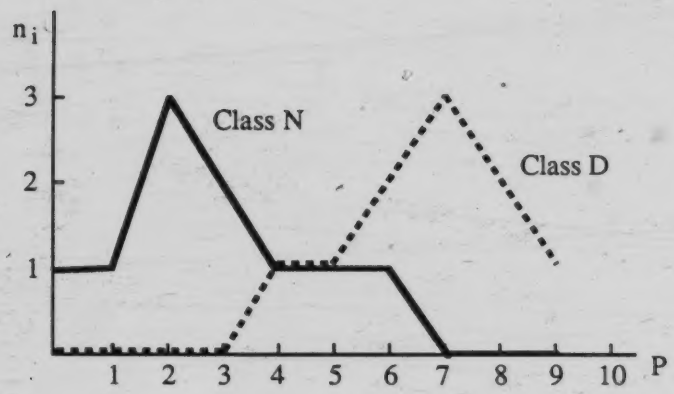


Figure 4.1 illustrates idea of pattern recognition. For a distribution of x for two classes, we classify the waveform belong to class 1 or class 2 depending on $g(x_1, x_2)>0$ or $g(x_1, x_2)<0$.

Class D	P	Class N	P
D1	5.8	**N1**	6.3
D2	7.4	**N2**	2.2
D3	8.6	**N3**	2.9
D4	8.5	**N4**	5.8
D5	6.6	**N5**	1.4
D6	9.3	**N6**	3.3
D7	7.4	**N7**	4.5
D8	6.7	**N8**	0.6
D9	4.1	**N9**	3.2
D10	7.2	**N10**	2.7

(a)



(b)

Figure 4.2 shows an example of one dimensional distribution. (a) is the values of function P for the data belong to different classes. (b) is a one dimensional distribution of the values of function P.

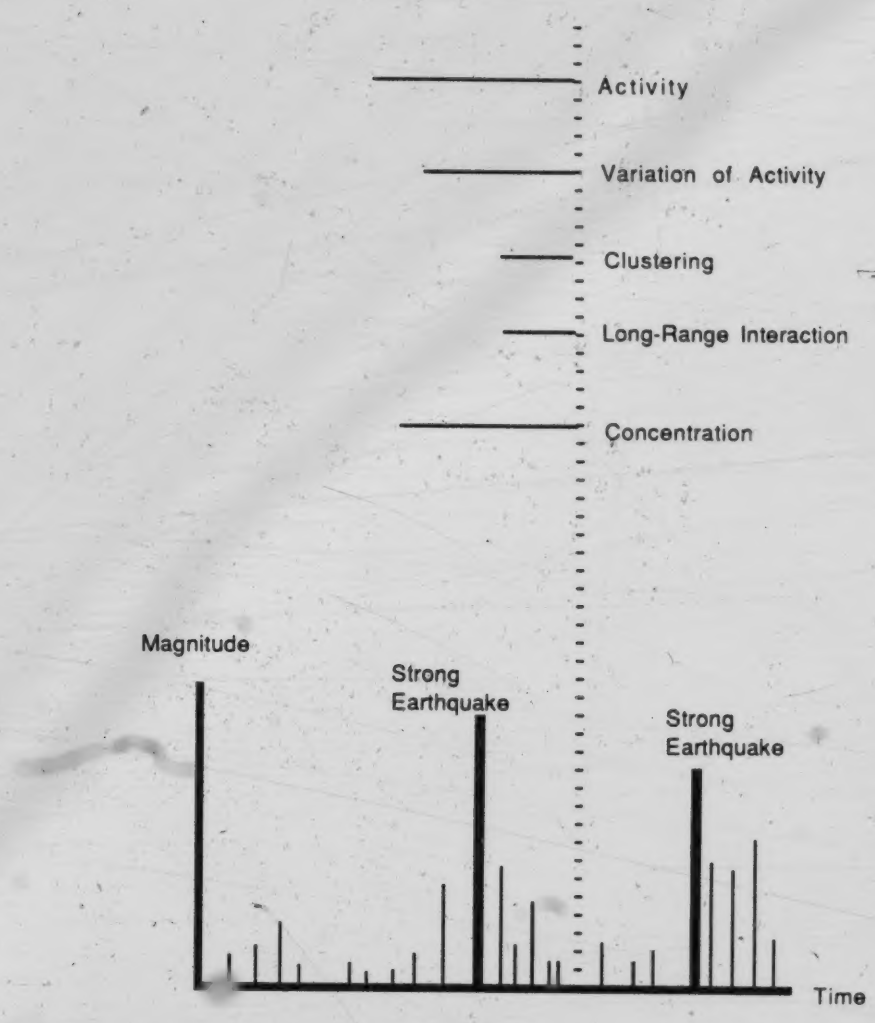


Figure 4.3 illustrates the idea of integral description for diagnosis of TIP. At each moment we look back in time and exam those characteristics of the earthquake sequence within the sliding time window, indicated by the horizontal lines, and determine if it belongs to a TIP (Adapted from Keilis-Borok, 1990).

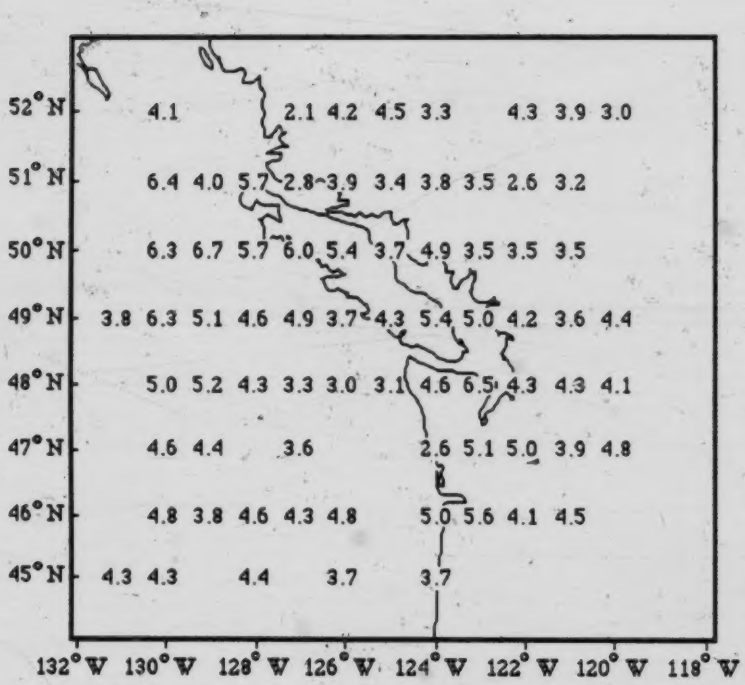


Figure 4.4. The largest earthquakes in 1^0x1^0 rectangles.

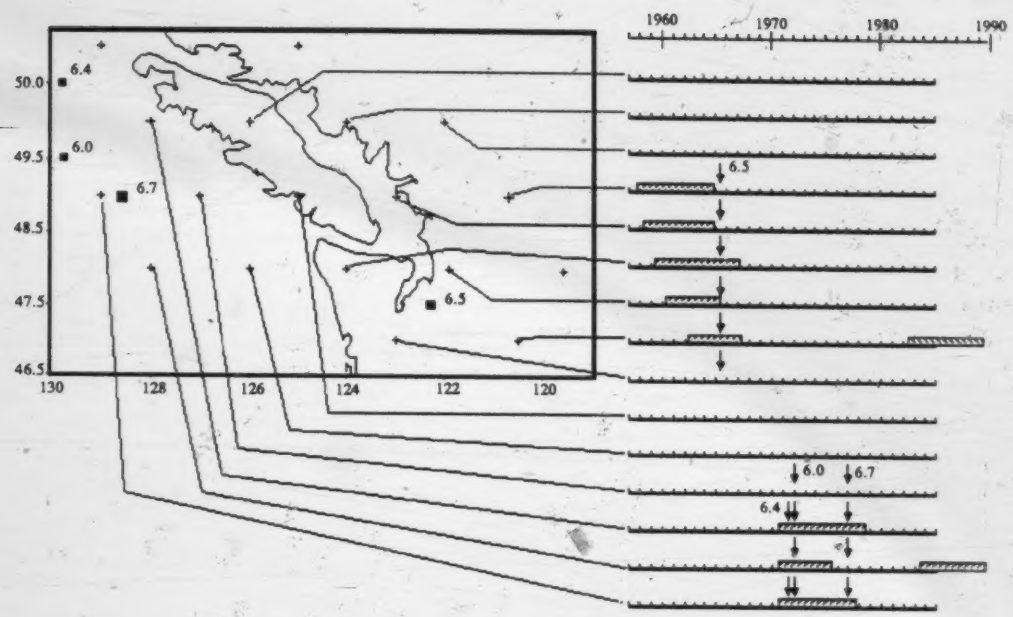


Figure 4.5. The centers of the regions analyzed with M8. A heavy line on the pointer to a location indicates the occurrence of a TIP in the time interval read from the scale in the upper right of the diagram. In this analysis we used a magnitude threshold of 6.5 and a start time in the catalog of 1952.

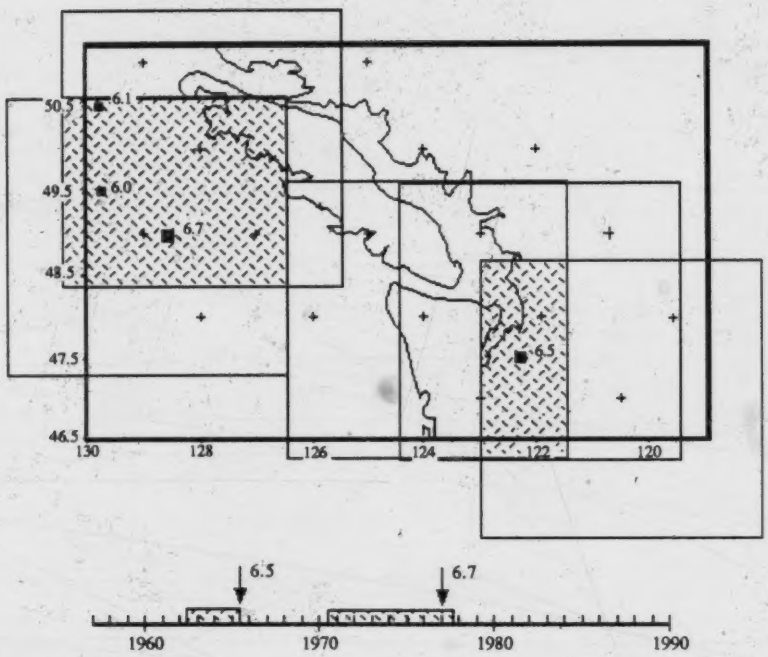


Figure 4.6. The regions covered by overlapping TIPs and one time intervals common to all TIPs for each event.

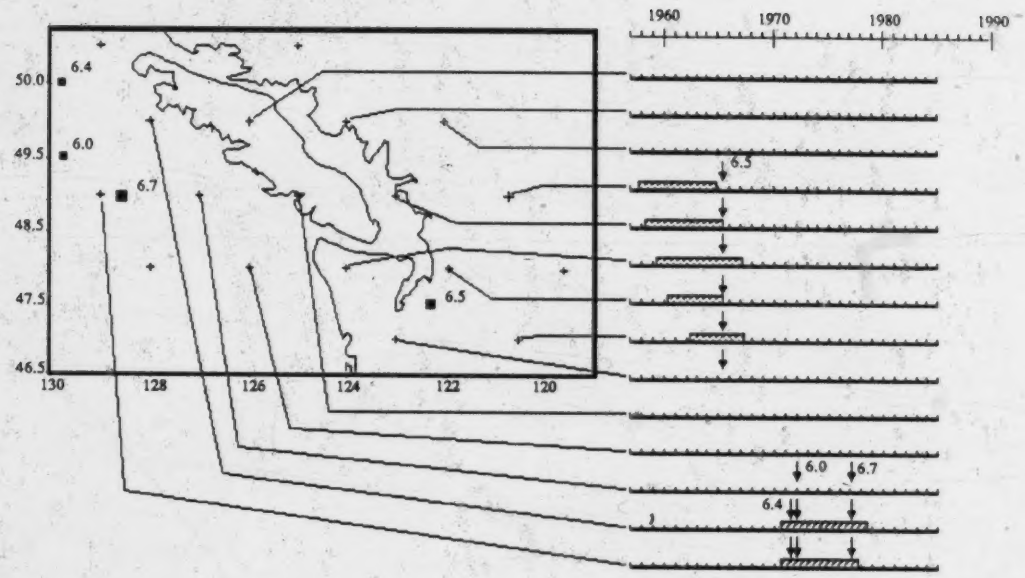


Figure 4.7. The results of the analysis using only events whose depths have been determined to be 25 km or less.

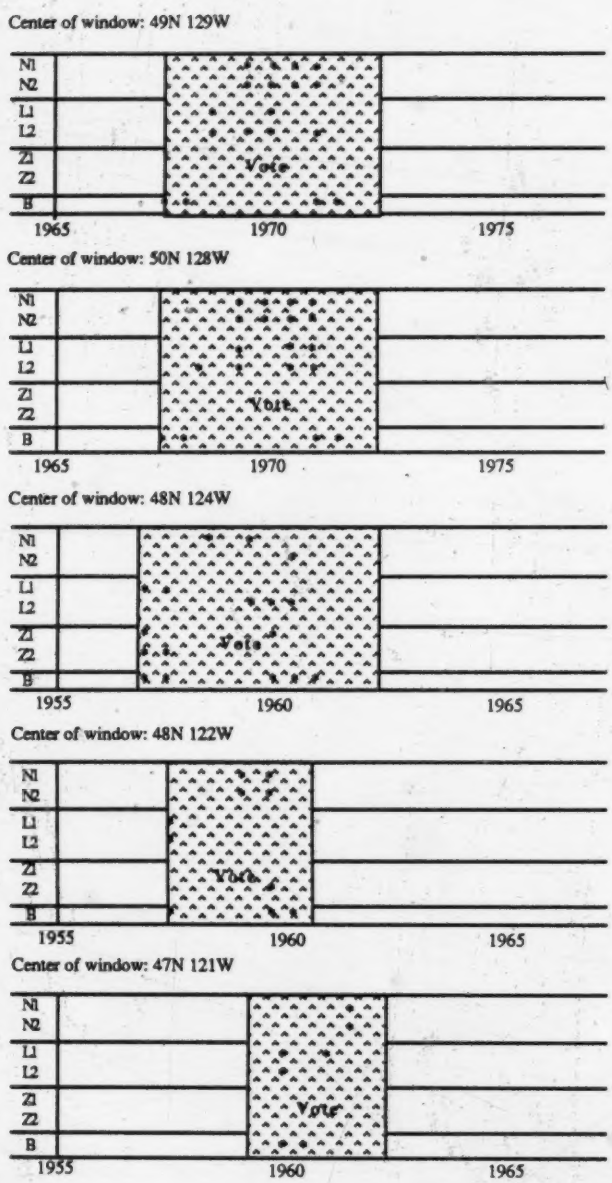


Figure 4.8. The part played by each function in establishing the TIP. The shaded area are the time intervals in which the activation was sufficient to establish a TIP (referred to here as a "vote").

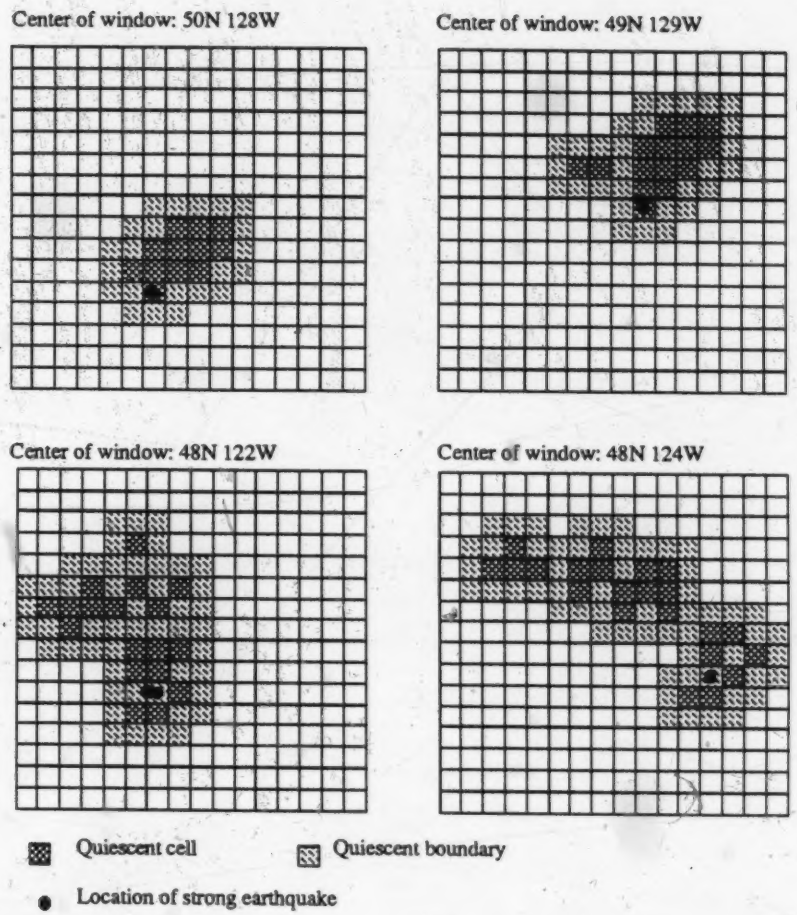

Center of window: 50N 128W
Center of window: 49N 129W
Center of window: 48N 122W
Center of window: 48N 124W
Quiescent cell
Quiescent boundary
Location of strong earthquake

Figure 4.9. Plots of activation and quiescence for the four large events in the catalog.

M	R (km)	T (days)
2.5	29	6
3.0	33	11.5
3.5	39	22
4.0	45	42
4.5	52	83
5.0	60	150
5.5	70	290
6.0	81	510
6.5	91	790
7.0	105	915
7.5	121	960
8.0	141	985

Table 4.1. If a mainshock has magnitude **M** or less, any event within **R** occurring within **T** days after the event and having a magnitude less than **M** is an aftershock.

			2.00	3.00		4.00		5.00		6.00		7.00	
			!-----!	-----!	-----!	-----!	-----!	-----!	-----!	-----!	-----!	-----!	-----!
1841	1	1	.	.	.	.	1	.	.	.	.	.	.
			.	.	.	.	1	.	.	.	.	.	.
1861	1	1	.	.	.	.	.	.	2	.	.	.	.
			.	.	.	.	.	.	.	.	.	.	1
1881	1	1	.	.	.	.	2	.	.	.	.	.	.
			.	.	.	.	2	1	4	.	.	.	.
1901	1	1	.	.	.	.	2	.	.	.	2	.	.
			.	.	.	.	14	.	2	3	2	1	1
1921	1	1	.	3	1	4	2	.	5	4	2	1	.
			.	.	.	.	12	1	4	4	1	2	.
1941	1	1	.	.	.	.	17	15	4	8	2	1	2
1951	1	1	42	.	22	10	5	6	.	1	.	.	.
			45	10	37	5	6	1	4	.	1	.	.
1955	1	1	84	37	22	9	4	.	1	.	1	.	.
			83	49	30	9	9	.	2	.	1	.	.
1959	1	1	61	28	24	14	3	7	1	.	1	.	.
			33	23	19	7	6	3	3	4	.	.	.
1963	1	1	19	17	18	15	24	8	5	.	.	.	.
			3	26	26	20	14	13	2	.	.	1	.
1967	1	1	24	39	52	14	15	4	2	.	.	.	.
			13	22	16	5	6	6	1	.	.	.	.
1971	1	1	3	9	15	4	13	9	4	2	3	.	.
			4	7	9	5	17	8	1	.	.	.	.
1975	1	1	40	42	25	15	9	11	7	.	.	1	.
			78	43	18	7	11	4	1	2	1	.	.
1979	1	1	67	42	23	55	152	71	13	.	.	.	.
			129	72	25	10	12	7	1	1	.	.	.
1983	1	1	77	48	24	2	3	5	.	.	.	.	.
			23	9	3	1	2	.	.	.	.	.	.
			!-----!	-----!	-----!	-----!	-----!	-----!	-----!	-----!	-----!	-----!	-----!

Table 4.2. shows a two dimensional time magnitude histogram of events in the catalog.

Year	Mon	Day	Hour	Min	Sec	Latitude	Longitude	Depth	Magnitude
1872	12	15	5	37		48.6	121.4		7.4
1917	12	23	15	48		50.0	128.0		6.5
1918	12	6	8	41	5.8	49.6	125.9		7.0
1926	11	1	1	39	18.0	48.8	128.5		6.6
1939	2	8	6	39	25.8	49.1	128.0		6.5
1946	6	23	17	13	25.8	49.8	125.3		7.3
1946	7	18	6	6	58.5	49.5	129.7		6.5
1965	4	29	15	28	44.0	47.4	122.3	59	6.5
1976	12	20	20	33	12.0	49.0	128.7	18	6.7

Table 4.3. The nine strongest earthquakes in the study region.

TIPs			Strong earthquake		Time of expectation*, months
Center of window		Time	date	M	
ϕ	λ				
49N	129W	1970,6,20 - 1977,6,20	1976,6,27 1971,3,13 1971,12,5	6.7 6.4 6.0	72
50N	128W	1970,6,20 - 1978,6,20	1976,6,27 1971,3,13 1971,12,5	6.7 6.4 6.0	72
48N	128W	1970,6,20 - 1975,6,20 1983,6,20 - 1989,6,20	1971,12,5 ?	6.0 ?	60 72
48N	124W	1959,4,29 - 1967,4,29	1965,4,29	6.5	72
49N	123W	1957,10,29 - 1964,10,29	–	–	84
48N	122W	1960,4,29 - 1965,4,29	1965,4,29	6.5	60
49N	121W	1957,10,29 - 1964,10,29	–	–	84
47N	121W	1962,4,29 - 1967,4,29 1982,4,29 - 1989,4,29	1965,4,29 ?	6.5 ?	36 84

* Time period from start time of a TIP to an strong event of $M>M_0$

Table 4.4. A summary of the TIPs for the Vancouver Island area.

Center of window: 50N 127W

			2.00		3.00		4.00		5.00		6.00		7.00	
1869	1	1	3	2	1	2	.	.	.	.	.	.	.	.
			.	7	3	.	1	2	.	.	.	.	.	.
1870	1	1	1	1	1	.	2	2	.	.	.	.	.	.
			1	3	3	3	1	1	.	.	.	.	.	.
1871	1	1	.	1	2	1	.	.	1	.	1	.	.	.
			1	1	2	.	1	.	1	1	1	.	.	.
1972	1	1	1	.	2	.	.	.	.	1	.	.	.	.
			.	1	.	1	.	.	.	.	.	.	.	.
1973	1	1	1	1	1	.	.	1	1	.	.	.	.	.
			1	2	1	.	1	.	.	.	.	.	.	.
1874	1	1	1	1	2	.	1	1	.	.	.	.	.	.
			.	2	2	1	3	1	.	.	.	.	.	.
1875	1	1	.	1	2	.	1	1	1	.	.	.	.	.
			4	6	3	.	.	2	.	.	.	.	.	.
1876	1	1	5	6	2	1	1	1	2	.	.	.	.	.
			4	1	1	1	2	1	1	.	.	.	1	.
1977	1	1	8	.	3	2	1	.	.	.	.	.	.	.
			8	5	2	.	1	.	.	.	.	.	.	.
1978	1	1	.	.	.	.	.	.	.	.	.	.	.	.

Center of window: 50N 127W

			2.00		3.00		4.00		5.00		6.00		7.00	
1858	1	1	13	7	5	.	1	.	.	.	.	.	.	.
			10	3	3	.	.	.	.	.	.	.	.	.
1859	1	1	9	4	2	.	.	.	.	.	.	.	.	.
			8	6	6	2	.	2	.	.	.	.	.	.
1860	1	1	8	1	2	.	.	.	.	.	.	.	.	.
			9	2	1	.	.	1	.	.	.	.	.	.
1961	1	1	6	7	1	.	.	.	.	.	.	.	.	.
			6	3	1	.	1	.	1	.	.	.	.	.
1962	1	1	5	1	1	.	.	.	.	.	.	.	.	.
			.	3	.	.	.	.	1	1	.	.	.	.
1863	1	1	4	2	.	.	.	.	.	.	.	.	.	.
			2	1	2	.	.	.	.	.	.	.	.	.
1864	1	1	.	1	1	.	.	.	.	.	.	.	.	.
			.	1	.	2	2	.	.	.	.	.	.	.
1865	1	1	1	.	.	.	.	.	.	.	.	.	1	.
			1	2	.	.	.	.	.	.	.	.	.	.
1966	1	1	.	.	.	.	.	.	.	.	.	.	.	.

Table 4.5. Magnitude-time histograms of events in regions centered on two large earthquakes.

Chapter 5: On modelling of the lithosphere dynamics

A model is generally needed to complement phenomenology of a physical system such as seismicity. In the field of seismicity building a model directly from the observational data is difficult due to limited coverage by the data both in time and space. But we have the alternatives of building a model from the fundamental principles of physics. Such a model is explanative. It allows us to understand more of the general behavior of the system and can generate a synthetic time series to complement the observable data.

How should we model the evolution of earthquakes? The discussion in the previous chapters and recent investigations of other (Horowitz, 1989; Huang and Turcotte, 1990) suggest that the geodynamical processes that generate seismicity in the lithosphere result in chaotic rather than random time series. It has been shown that the lithosphere dynamics is characterized with a low dimensional attractor. This indicates a deterministic model that describes the dynamical evolution of the lithosphere is possible. Lithosphere dynamics should appropriately be described in a deterministic way. The evolution of earthquakes should be studied together with the entire dynamics of the lithosphere since they are the integral part of the lithosphere dynamics.

Our understanding of the lithosphere dynamics results, by and large, from mechanical models for a single fault isolated from the rest of the lithosphere. The insight into lithosphere dynamics we draw from these models is limited only to a certain spatial scale. Any extension of such an approach to general lithosphere dynamics could be wrong, since the lithosphere is characterized with hierarchical structure of volumes. Little study has been done to characterize the integral dynamic evolution of the lithosphere, but despite our lack of physical understanding, an analogy can be drawn between some of the phenomenology of lithosphere dynamics and the general behavior of other non-linear dynamic systems.

For example, Kagan and Knopoff (1980, 1981) have shown that earthquake fault zones have three dimensional structure that has a fractal distribution of size. The fractality implies that the underlying dynamic field, such as stress and strain field, governing the occurrence of earthquakes does not have an intrinsic length scale. This is reminiscent of a

very similar phenomenon related to the velocity field of a turbulent flow, where the flow is governed by the Navier-Stokes equation that is scale invariant. When the Reynolds number becomes large, the flow will change from a laminar flow to a turbulent flow with eddies on all scales. The length scale imposed on the system due to the boundary effect will become weaker when the flow is away from the boundary. The symmetry of the equation is partly recovered. I believe there is a analogy between the velocity field of a turbulent flow and the strain or stress field of the lithosphere. The dynamics of the lithosphere might be similar, in some aspects, to the development of a turbulent flow. Through the interactions between fault zones, or their related stress and strain field, the fault zone will develope to a structure with a fractal distribution of size, that is not affected by the length scale imposed on the system by the regional tectonics.

It is my hope that the dynamics of the lithosphere is a system of this kind. There might exists an as yet unidentified group of non-linear differential equations that describe the dynamics of the lithosphere. In such a system only the values of some state variables will distinguish fault zones from intact areas of the lithosphere, and time evolution of these state variables is governed by the differential equations. Just as in the scenario of a turbulent flow, the strain or the state fields will have regular behavior under small deformation rate, and become irregular under the large one. In the chaotic regime the processes will generate some spatial structures which may display a self-similarity. The present fractal structure of the lithosphere is not necessarily an inherent property of the lithosphere. Since the occurrence of earthquakes is a process of faulting, it is designated in this model as the weakening in the fault zone corresponding to the change of a state variable, and therefore causes the strain or stress field change in the surrounding medium. Since the fault is inherent in the model, seismicity must be studied as an integral part of the development of the entire dynamics of the lithosphere.

I believe the chaos established in the lithosphere is a result of coherent interaction between fault systems. The block model of the lithosphere (Gabrielov, et al., 1986), which assumes that the lithosphere consists of many blocks and earthquakes occur due to the shear deformation along the block boundaries, suggests that the interactions between blocks dominate the dynamics of the lithosphere. It also shows a scaling in the magnitude

and frequency relation for the earthquakes simulated in the model. Theoretical treatment of the medium like this is the major difficulty involved in the modelling the dynamics of the lithosphere. The lithosphere is basically full of fractures, while the continuum assumption is required in most theories dealing with mechanical deformation of a medium.

In this chapter, I attempt to formulate the stress and strain field of a blocky lithosphere by introducing two state variables. I relate these state variables to the internal structure of the block arrangement following the rationale of the statistical mechanics (Landau and Lifshitz, 1958; Pathria, 1972; Gong, 1982). I expect that a certain extent of the macroscopic response of the internal structure of the lithosphere could be reflected in these state variables, and that it can be used to characterize a fault zone. This avoids treating the fault as a special boundary with constraints artificially imposed on it.

5.1 Mesoscopic structure of the lithosphere

The lithosphere is different when it is measured with different spatial scales (Figure 5.1). With the largest measuring scale, the macroscopic scale, detailed structures of the lithosphere are smeared, and the lithosphere behaves like a continuum. At an intermediate scale called here the mesoscopic scale, the lithosphere consists of many small solid phase blocks. It is a discontinuous material at this scale. The interfaces between blocks play a significant role in the mechanical behavior of the continuous phase, particularly when the the structural arrangement breaks down. At this resolution each individual solid block can be treated as a continuous body to which the theory of solid mechanics applies. The smallest measuring scale, the microscopic scale, applies to the atomic phase. With this resolution we can look into the lattice structure of the solid phase, and explained the viscoelastic and plastic behavior of the solid phase in terms of lattice structure and its defects. Axelrad (1984) first introduced the concept of the mesodomain in the probabilistic mechanics. He argued its existence within the medium form the link between the microscopic and macroscopic description of the material behavior.

Most approaches to modelling the mechanical behavior of the lithosphere are applied to the continuous phase of the lithosphere, and ignore the presence of the

mesoscopic structure in the material. A continuum theory, such as elasticity or viscoelasticity, is generally adopted as a mathematical tool to describe the mechanical state and its time evolution of the system. But such approaches fail when we deal with fracture processes or earthquake dynamics. In this circumstances the lithosphere can not be approximated as a "fine" continuous phase without any internal structure. Structural arrangement changes in the mesodomain will play a part in the macroscopic properties of the continuous phase. Therefore, besides considering the response of various classes of the continuous phase, we should equally consider the effects that are due to the inherent geometrical and physical properties of the structural elements in mesodomain.

The complete understanding of the dynamics in mesodomain can be obtained, conceptually, by writing down the elasticity or viscoelasticity equations for each solid phase block and solving the boundary value problem, but that is mathematically impossible due to the complexity of the system. Shi (1988) proposed a reasonable approach to solve the system of this kind in his suggested method Discontinuous Deformation Analysis. In his formulation, he describes each solid phase block by the coordinates of its center of mass $\mathbf{r}_c$, the rotation angles θ, and the deformation $\mathbf{e}_{ij}$. The dynamic evolution of each block is described by their corresponding time derivatives $\dot{\mathbf{r}}_c$, $\dot{\theta}$, and $\dot{e}_{ij}$ (Figure 5.2). The mechanical properties of the blocks are described by the material parameters such as elastic constants and viscosity coefficient. The strain $\mathbf{e}_{ij}$ is taken as a basic variable here, it accounts for the deformation of the block under the first order approximation. The total potential energy of the system is designated Π_p, which is the summation over all the potential energy sources, individual stresses and forces. With the quantities defined above, the total potential energy Π_p can be expressed as a function of $\mathbf{r}_c$, θ and $\mathbf{e}_{ij}$ of all the elements considered. For a specific configuration of solid phase blocks and a given boundary condition, only those among the admissible states that minimize the total potential energy Π_p are stable and physically possible. Discontinuous Deformation Analysis works well for a system of a reasonably low number of degree of freedoms. When the number of degree of freedom however increases to the level required in the lithosphere, the calculations needed to solve the problem will be beyond the capacity of any modern

computer.

In the problems of the dynamics of the lithosphere we are not interested in the mechanical behavior of each solid phase blocks but rather in their effects on the mechanical behavior of the continuous phase in the macrodomain. I suggest here that the macroscopic properties of the lithosphere are connected with the mesoscopic composition of the system. Also the number of macroscopic quantities used to describe the continuous phase in the macrodomain is generally much less than the number of degree of freedoms of the system in mesodomain. Macroscopic properties of the lithosphere do not depend on what each individual solid phase block is doing, rather depend only on some average behavior of all the blocks. A given macroscopic state may correspond to numerous structural configurations in the mesodomain which are compatible with a given macroscopic specification. Thus we could take advantage of this to study the problems of predicting the average behavior of many blocks.

5.2 Phenomenological formulation in mesodomain

A system with a great number of degree of freedoms generally has many stable states corresponding to different structural configurations. The set of all possible stable states of the system forms a sample space over which the statistics of many solid phase blocks can be examined. Under the internal interactions or perturbation from outside, the system will change from one state to another. Some dynamic systems may experience all their possible states during a period of measurement of their macroscopic quantities. Therefore, the quantities measured must be considered as time averaged values, which are hopefully the same as those averaged over a statistical ensemble, a collection of a large number of identical systems. Since the time needed to take a measurement of a physical quantity over a selected sample of the lithosphere is much shorter than the time during which the system experiences all the possible states, the mesoscopic structure of the selected system therefore can be considered fixed, at least for the duration of the measurement. Yet, for a collection of samples of the lithosphere that are compatible with a given macroscopic specification, the measurement of a physical quantity are generally

different for different samples. The average value of a measured observable, which is taken over the statistical ensemble, should therefore be taken as the best obtainable estimate of the actual value for the sample. Such a value is the most probable value of the ensemble, it corresponds to many more mesoscopic states than other values do. Therefore, it is reasonable to expect the lithosphere is in such a state, and the fluctuation is small.

Consider a large collection of isolated identical systems, which is usually called a microcanonical ensemble, each system exists in a specific mesostate. Suppose we try to measure a physical quantity A, we will find each system will give a different value of A. By counting how many systems have the value of A within the interval $A+dA$, we will obtain the probability distribution $\rho(A)$. The average value of A is therefore given as

$$\overline{A} = \int_{-\infty}^{\infty} \rho(A)A \, dA$$

The expectation $\overline{A}$ is the best guess of the value we will obtain prior to an observation. We assume that among all the possible distributions of $\rho(A)$, there is one most probable distribution which corresponds to the reality of the nature. The distribution can reasonably be expected to have a sharp peak around the expectation $\overline{A}$, which is therefore the most probable value of A. The states at which the system will yield a value of A around $\overline{A}$ are the most probable states of a natural system. Therefore, the situation will become much easier if we study only those physical quantities associated with the most probable states, and believe that they are the reasonable representation of the nature of the system.

In thermodynamics a special state we are interested in is the thermal equilibrium state, which is the state a thermalized system tends to reach. We use the word equilibrium in thermodynamics when the system reaches the dynamical equilibrium, but the situation discussed here is different. Therefore, we emphasize the word "the most probable" to indicate the state we will most probably encounter in the nature.

Consider a microcanonical ensemble with N systems, each system has Ω number of stable states under a given deformation, where Ω is much smaller than N. If there are N_i systems in the state Ω_i, the probability of finding the system in state $\Omega_i$is

$$P_i = \frac{N_i}{N}$$

With the above probability distribution, the entropy of the system is defined as

$$S = - \sum_i P_i \log P_i$$

The sum is over all possible states. Shannon introduced this function and argued that it measures the amount of information in any message (Shannon and Weaver, 1962). For the system considered here, entropy can be said to measure our ignorance of the well defined mesoscopic configuration of the system, if all we know is the probability distribution.

Maximum entropy describes maximum ignorance when we know absolutely nothing about the system. Therefore, the distribution obtained by maximizing the entropy will give the most reasonable assumption about the distribution of the system one could give without having any knowledge of the system. It is also the most probable distribution the system can have if there is no constraint imposed on the systems. It is not surprising that a uniform distribution in which all the states of the system are equally probable gives the maximum entropy.

If we gain some knowledge about the system, the probability distribution will change, and so will the entropy. For the system considered above, we assume the system has energy E_i when it is in the state Ω_i. If we make an observation of the energy E, the best we can do is to infer that it represents the expectation energy $\overline{E}$. To generate the appropriate probability distribution, we maximize the entropy with the following constraints:

$$\sum_i P_i = 1$$
$$\sum_i P_i E_i = \overline{E} \qquad (5.1)$$

By introducing two Lagrangian multipliers α and β, we are led to the Boltzmann distribution:

$$P_i = e^{-\alpha - \beta E_i} \qquad (5.2)$$

of statistical mechanics. It is the most probable distribution of the system under the above constraints. Introducing the notation $\Omega(E_i)$ which represents the number of states with the same energy E_i, we have

$$P_{E_i} = \frac{1}{Z}\Omega(E_i)e^{-\beta E_i}$$

where P_{E_i}, the probability that the system is in a state of energy E_i, depends both on the energy E_i and the number of states $\Omega(E_i)$ at this energy. The term $e^{-\beta E_i}$ indicates that the chance of finding a system in a high energy state is small, on other hand since $\Omega(E_i)$ generally increases when the energy increases, this will increase the probability. In the above expression α is included in Z, which is known as partition function in statistical mechanics.

$$Z = \sum_{E_i} \Omega(E_i)e^{-\beta E_i}$$

The partition function is a controlling function which determines the average values of physical quantities.

If the system also depends on parameters $a^1, a^2, \ldots a^s$, the most probable distribution is obtained by maximizing the entropy with the following constraints as well as (5.1),

$$\sum_i P_i a_i^m = \overline{a^m} \qquad\qquad m=1, 2, \ldots s$$

The resulting distribution is (Appendix C)

$$P_i = e^{-\alpha-\beta(E_i-\sum_m \lambda_m a_i^m)}$$

or expressed in an equivalent form when the number of states corresponding to the energy E_i and parameters $a^1, a^2, \ldots a^s$ is denoted by $\Omega(E_i,\{a_i^m\})$

$$P_{E_i,\{a_i^m\}} = \frac{1}{Z}\Omega(E_i,\{a_i^m\})e^{-\beta(E_i-\sum_m \lambda_m a_i^m)}$$

where λ_m are partial derivatives of the energy,

$$\lambda_m = \left(\frac{\partial E_i}{\partial a_i^m}\right)_S \qquad m=1,2,...s$$

For a mechanical system, we replace the parameters a_i^m by the strain e_{ij}. When the strain e_{ij} changes, the energy of the system also will change, and the partial derivatives

$$\frac{\partial E}{\partial e_{ij}} = \sigma_{ij}$$

indicate the stress acting on the system.

In the above formulation we have two quantities, S and β, associated with the most probable distribution. They generally depend on the other macroscopic parameters, such as strain and stress, that control the system. We have a form of a very well known relation (Appendix C)

$$\frac{1}{\beta} dS = d\bar{E} - \sigma_{ij} de_{ij}$$

This is the first law of thermodynamics in mesodomain. It indicates that the total energy stored in the system is composed of two parts: one is due to the work done on the system that is the energy associated with macroscopic deformation, another is an energy associated with random arrangement of the solid phase blocks.

This formulation can be applied to any physical system with a large number of degree of freedoms, when its macroscopic properties are associated with a distribution. It is not specific to the mechanical system we consider. The argument is that the lithosphere can be considered a system of this kind. Therefore, besides the strain or stress that are generally used to describe the mechanical state of the lithosphere, two more state variables, S and β, should be used to describe the state of internal structure of the lithosphere. The rationale behind this consideration is the assumption that among all the possible distributions of the mesoscopic states of the system there is a most probable one, and all the questions about the mechanical properties of the system are only addressed to the system at such a state.

The mutual relationship between state variables, here is stress, strain and β or S, is the equation of state. By assuming a geometry and mechanical properties for the solid phase blocks, one can, in principle, derive the equation of state, but it is mathematically very difficult for realistic cases. It may be appropriate in practice to obtain such an equation of state through an intelligent guess. The implicit assumption embodied in the possibility of making an useful guess about the equation of state is that the effects of geometry of the solid phase blocks is a high order correction to the equation of state. The rationale behind this consideration is the analogy to the case of the classical gas, where the effects of the molecular structure of the gas is added as an high order correction to the equation of state for the ideal gas.

5.3 Equation of state for a one dimensional material

A one dimensional model is a trivial example of such an approach, but I discuss it due to its simplicity in mathematics. It allows an explicit evaluation of the partition function Z, from which the average values of macroscopic quantities of the model can be calculated. Although the model is naive, it is instructive as a way of demonstrating how the mechanical properties of a material might depend on those parameters introduced above, and how to proceed in more realistic cases.

Consider a system of N elastic blocks of side length a and b. There are two possible states for a block in a chain which forms a one dimensional material. Under the constant force F acting on the both ends of the material, each block will contain a certain amount of strain energy corresponding to which state it is in. When the block is in the state a, (Figure 5.3), the strain of the block is

$$e_a = \frac{F}{b\,Y}$$

where Y is the elastic constant of the block. The corresponding strain energy stored in the block is

$$E_a = \frac{1}{2} \sigma e_a V = \frac{1}{2} \frac{F^2 a}{Yb}$$

where V is the volume of the block. The same reasoning applies when the block is in the state b, the strain energy stored in the block in this case is

$$E_b = \frac{1}{2}\sigma e_b V = \frac{1}{2}\frac{F^2 b}{Ya}$$

Assume that there are N_a blocks in the state a and N_b blocks in the state b. The total strain energy E_i and length L of the system will depend on the N_a and N_b in the following way

$$E_i(L) = N_a E_a + N_b E_b$$

$$L = N_a a + N_b b$$

For the given distribution N_a and N_b corresponding to the system of energy E_i and length L, the total number of ways in which the process of distribution can be carried out is

$$\Omega(E_i, L) = \frac{N!}{N_a! N_b!}$$

Substitute the above relations into the partition function Z, where we replace λ_m by the force $-F$ and a_i^m by the length L,

$$Z = \sum_{E_i,\, L} \Omega(E_i, L) \exp\{-\beta E_i(L) + \beta F L\}$$

$$= \sum_{N_a} C_{N_a}^{N} (e^{\beta(F a - E_a)})^{N_a} (e^{\beta(F b - E_b)})^{N - N_a}$$

$$= [e^{\beta(F a - E_a)} + e^{\beta(F b - E_b)}]^N$$

The equation of state of the material can be obtained by differentiating the partition function with respect to the βF,

$$L = \frac{\partial \log Z}{\partial \beta F} = N \frac{a e^{\beta(F a - E_a)} + b e^{\beta(F b - E_b)}}{e^{\beta(F a - E_a)} + e^{\beta(F b - E_b)}}$$

Replace E_a and E_b by $F^2 b/2Ya$ and $F^2 a/2Yb$ respectively, we have

$$L = N\frac{a + b\, e^{\beta F \gamma}}{1 + e^{\beta F \gamma}}$$

where $\gamma = (b - a) - \frac{F}{2Y}\left(\frac{b}{a} - \frac{a}{b}\right)$. a and b are taken as constants for their zeroth order approximation. The mutual relation among L, β and F is plotted in a 3D plot (Figure 5.4a). The quantities L, β and F shown in the plot are non-dimensional quantities, they are normalized by the factors Na, $1/Ya^2$ and Ya respectively. The ratio between b and a is chosen as 1/2.

The relation between L and F is highly nonlinear. As β increase, L changes very rapidly with F. The derivative of L with F is plotted in figure 5.4b. The deep notch in the figure indicates a sudden weakening of the material as the force increases. The state of the system at this point is very unstable, the sudden transition from one state to another can be explained as an earthquake.

The physical meaning of β is clear in this model. The smaller the value of β, the more random is the arrangement of the blocks. When the value of β tends to zero, about half of the blocks are in state a and half are in state b, this is the most random state. When the value of β tends to infinity, most of the blocks will be in state b.

5.4 Discussion

Introducing two extra state variables, which are not independent, can account for the effect of complicated blocky structure in mesodomain on larger scale in macrodomain of the lithosphere. In principle the equation of state can be derived for the material that is composed of solid blocks of certain geometry and mechanical properties. It gives a relation among stress, strain and one state variables. This implies that the state variable is uniquely determined by the instantaneous values of the stress and strain. If the state variable is kept constant there will be a unique relationship between the stress and strain. For the fast changes of stress and strain, the state is expected to be unchanged. These implications indicate the existence of a family of curves on stress and strain plane that cannot intersect.

This provides a necessary condition for the legitimacy of above consideration that can be verified by exhaustive experiments (Ruina, 1983).

Since the strain energy in a sample can be measured, there will be a unique relationship between the energy and strain if we keep stress constant. Discontinuous Deformation Analysis (Shi, 1988) provides a numerical way of testing this relationship.

The arguments refer to static problem. The description of the system is incomplete unless we include the equations that describe the time evolution of the state variables. The dynamical evolution of the system may be determined from the investigation of dynamic interaction of the solid blocks in mesodomain and the interaction between the solid block phase and the atomic phase in microdomain. Such investigations may encompasses all the complexities of the system, and are not discussed here.

In the one dimensional model the occurrence of an earthquake corresponds to a sudden transition from one state to another. A two dimensional model can behave very differently, since such a model allows both normal and shear stress. We can speculate that a fault in this case is a weak zone where entropy is probably high; the blocks are more randomly organized therefore allow shear displacement to occur more easily (Figure 5.5). When the entropy is low, the blocks are well organized and therefore they are interlocked with each other. The material in this case has a high shear strength. The occurrence of an earthquake is probably associated with a sudden increase of the entropy, or weakening of the material in the fault zone. After the earthquake, the blocks start rheological deformation under the tectonic stress, and tend to become more organized. The entropy therefore will decrease and this can be considered as healing of a fault. Since rheological phenomena are always associated with heat generation, they produce entropy in atomic phase. From this we see that healing is a result of interaction between the solid block phase in the mesodomain and the atomic phase in microdomain.

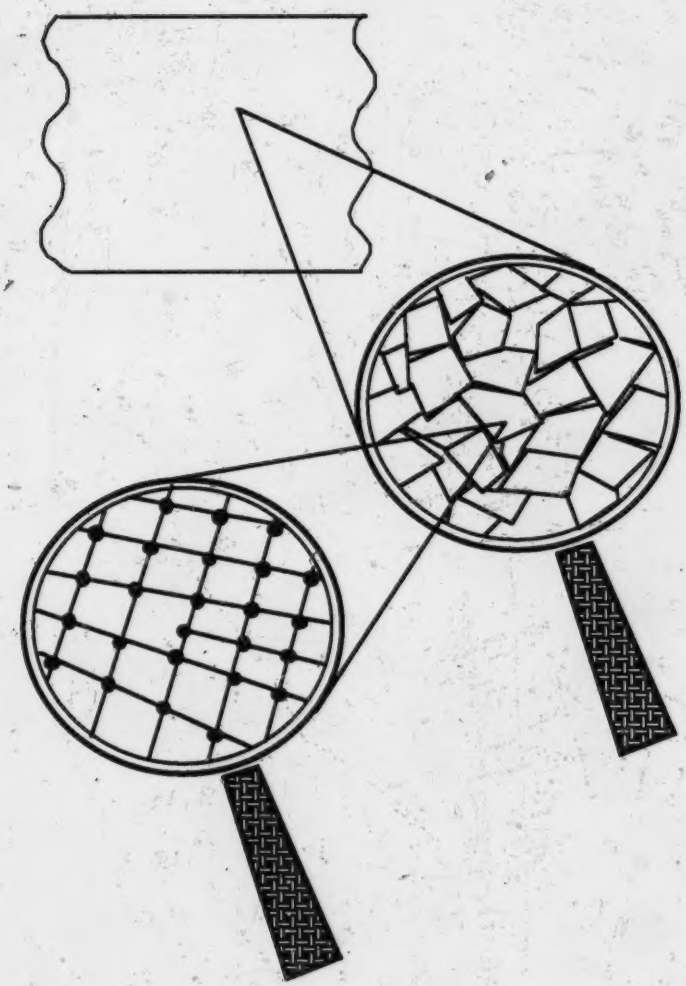


Figure 5.1 illustrates the idea of dividing the lithosphere into three phases according to three measuring scales.

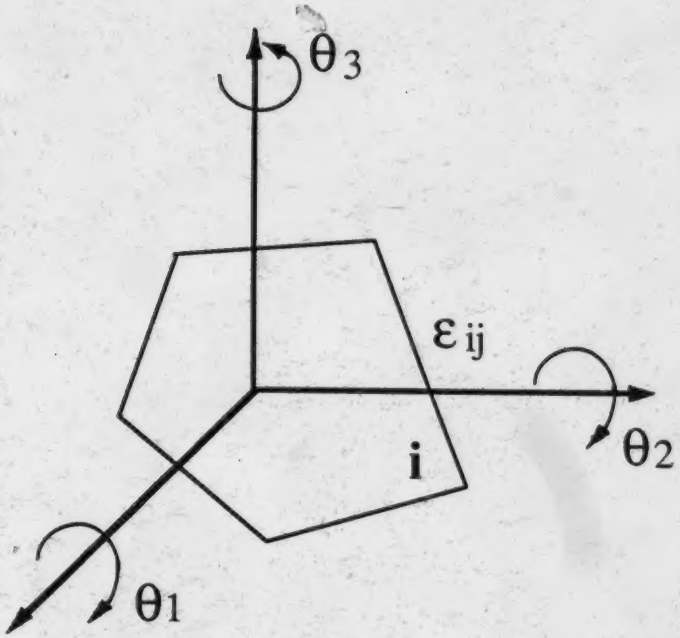


Figure 5.2 shows one block of a block system. It is described by its position, rotation and deformation.

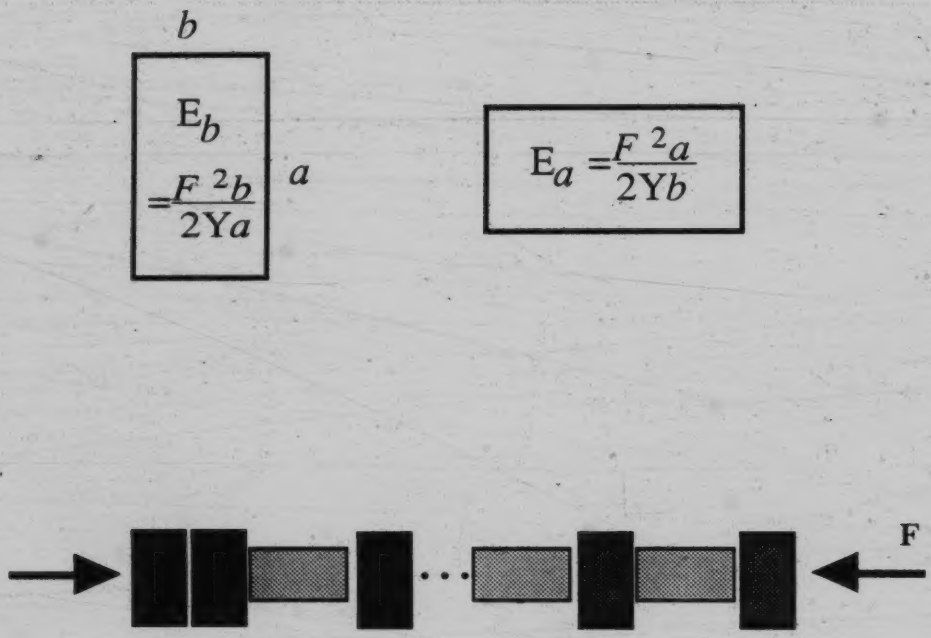


Figure 5.3 illustrates a one dimensional block model which consists of a chain of blocks in different states.

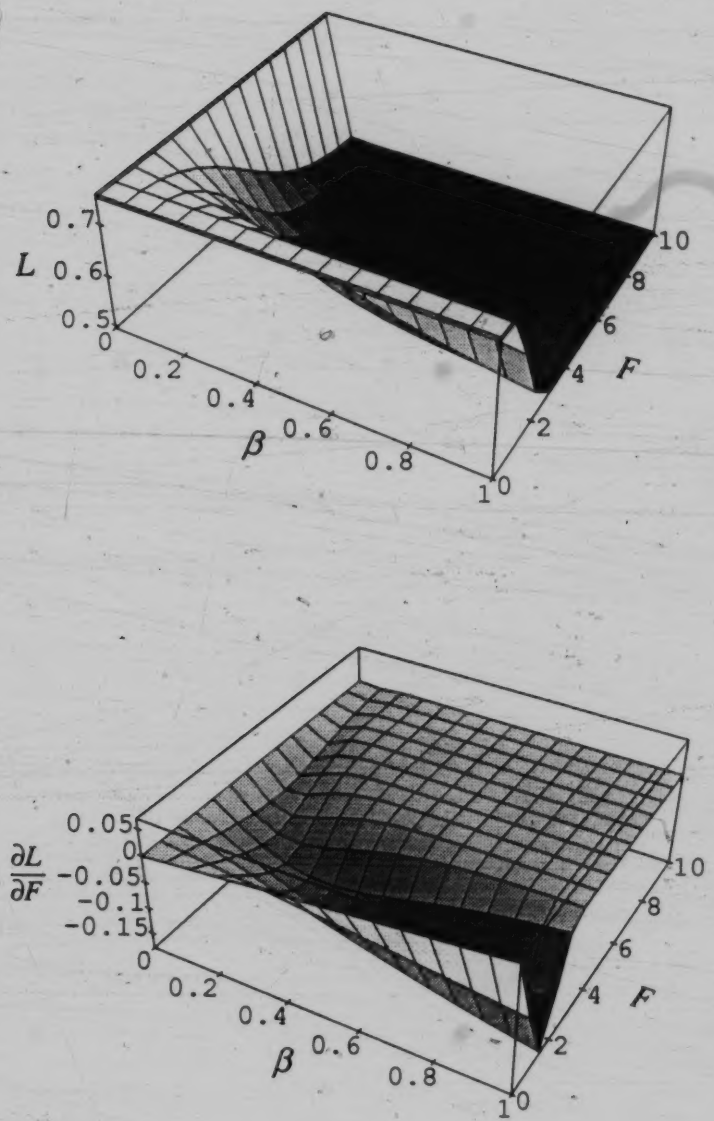


Figure 5.4 (a) shows the mutual relations among L, β and F. (b) shows the relation among the derivative of L with respect to F, β and F.

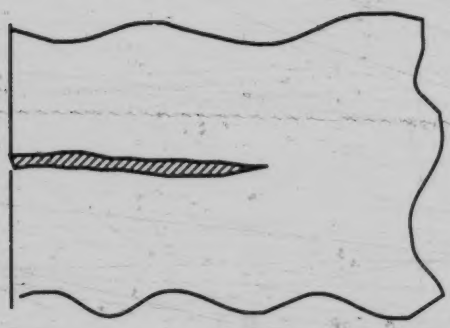


Figure 5.5 sketches a two dimensional model, in which a fault is described as a weak zone where the blocks are randomly organized. It corresponds to a state of high entropy.

Concluding Remarks

This thesis has introduced the use of nonlinear dynamics methods for the study of lithosphere dynamics and earthquake prediction. The primary objective of this study is to show that it is more appropriate to treat the dynamics of the lithosphere as a deterministic nonlinear dynamic system. I believe that better understanding of occurrence of earthquakes can result from investigations of the nonlinearity of the lithosphere dynamics. The results are not exhaustive and further investigations in other areas are needed to extend the results.

I consider the dynamics of the lithosphere to be represented by a nonlinear deterministic system. Earthquakes are an integral feature of lithosphere dynamics since they correspond to a collection of Poincare points in the phase space of the lithosphere. The assumption behind this consideration is that the lithosphere dynamics can be described by an as yet unknown group of differential equations. This leads to the result that the attractor of the lithosphere dynamics can be reconstructed from the earthquake catalogs. Therefore, all the properties of the attractor obtained from analyzing the earthquake catalogs are intrinsic to the lithosphere dynamics. For example, the low dimensionality of the attractor, between 3 and 4, found from the analysis of the earthquake catalog of the west coast of Canada suggests lithosphere dynamics is chaotic rather than random. Such an analysis of other geodynamic measurements should yield the same dimensionality. The upper limit of time for earthquake prediction, found to be 20 years on the west coast of Canada, is a fundamental characteristic time of such dynamics. It is not only a limit for earthquake prediction but also a limit of the predictability of any other geodynamic quantities.

The determination of the physical model of the lithosphere and its differential equations is still in its infancy. It has produced few results to support this conceptual frame work of the lithosphere dynamics, but such considerations are meaningful because they predict some results that can be tested experimentally. For example, it predicts that other geodynamic measurements of the lithosphere are topologically equivalent to earthquake catalogs. Investigation of the equivalence between earthquake catalogs and other geodynamical measurements, such as strain meter measurements, can be used to test

whether such consideration is legitimate. At the present stage of this study it is reasonable to believe that lithosphere dynamics is such a nonlinear system for it displays a similar phenomenology to other nonlinear dynamic systems, such as fluid dynamics.

The argument made here implies a new perspective for earthquake prediction. We can retrieve enough information from an ideal earthquake catalog to predict the future. More geodynamic measurements will not result in additional information about the dynamics of the lithosphere. Still, when we deal with the imperfect earthquake catalog we encounter in practice, many difficulties arise. These difficulties may be overcome by analyzing other geodynamic measurements. It is also possible that the pattern recognition algorithm designated to analyze earthquake catalogs can be extended to other geodynamic measurements. It is intuitively obvious that more measurements of different quantities will yield greater predictability.

The low dimensionality found in the earthquake catalog suggests that earthquake flow (Keilis-Borok, 1990) only occupies a low dimensional subspace. With such a low dimensionality, between 3 and 4, it seems possible to build a model directly from the data for some areas with good seismic records (Masdagli, 1989). For noisy data that we will inevitably encounter when we deal with geodynamic measurements, there is still need for better numerical methods to build a successful model which is not discussed in this thesis.

A physical model may not directly relate to the problem of earthquake prediction, but it helps explain why lithosphere dynamics behaves as it does. It suggests generalizations beyond earthquake prediction for a particular region, and allows us to relate the dynamics of the lithosphere from one area to another, allows predictions about how the lithosphere dynamics will vary with system parameters. The preliminary attempt toward this objective I made is to introduce two state variables in addition to stress and strain to describe the lithosphere state. These state variables reflect, to a certain extent, the complexity and interaction in sub-fault systems. The result from a one dimensional model shows that the constitutive relation is very nonlinear, and catastrophic failure could happen at the critical point.

References

Aki, K. and P. G. Richards (1980). *Quantitative Seismology, Theory and Methods*, W. H. Freeman and Company, San Francisco.

Allegre, C. J., J. L. Le Mouel and A. Provost (1982). Scaling rules in rock fracture and possible implications for earthquake prediction, *Nature*, 297, 47-49.

Ando, M. (1975). Source mechanisms and tectonic significance of historical earthquakes along the Nankai trough, Japan. *Tectonophysics*, 27, 119-140.

Andrews, D. J. (1980). A stochastic fault model I, static case, *J. Geophys. Res.*, 85, 3867-3877.

Andrews, D. J. (1981). A stochastic fault model II, time-dependent case,. *J. Geophys. Res.*, 86, 10821-10834.

Axelrad, D. R. (1984). *Probabilistic Mechanics of Discrete Media*, Pergamon Press, New York.

Basham, P. W., D. H. Weichert, F. M. Anglin and M. J. Berry (1982). New probabilistic strong seismic ground motion maps of Canada: a compilation of earthquake source zones, method and results, *Earth Phys. Branch. Open File* 82-33.

Basham, P. W., D. H. Weichert, F. M. Anglin and M. J. Berry (1985). New probabilistic strong seismic ground motion maps of Canada, *Bull. Seismol. Soc. Am.*, 75,563-595.

Bennet, J. G. (1936). Brocken coal, *J. Inst. Fuel*, 10, 22-39.

Broomhead, D. S. and G. P. King (1986). Extracting qualitative dynamics from experimental data, *Physica D*, 20, 217-236.

Brown, D., Q. Li, E. Nyland, and D.H. Weichert (1989). Premonitory seismicity pattern near Vancouver Island, Canada, *Tectonophysics*, 167, 299-312.

Burridge, R. and L. Knopoff (1967). Model and theoretical seismicity, *Bull. Seismol. Soc. Am.*, 57, 342-371.

Byerlee, J. D. and W. F. Brace (1968). Stick-slip sliding and earthquakes effects of rock type, pressure, strain rate and stiffness, *J. Geophys. Res.*, 73, 6031-6037.

Carlson, J. M. and J. S. Langer (1989). Mechanical model of an earthquake fault, *Phys. Rev. A*, 40, 6470-6484.

Casdagli, M (1989). Nonlinear prediction of chaotic time series, *Physica D*, 35, 335 356.

Crutchfield, J. P., J. D. Farmer, N. H. Packard and R. S. Shaw (1986). Chaos, *Sci. Am.*, 255, 46-57.

Dieterich, J. H. (1978). Time-dependent friction and the mechanics of stick-slip, *Pure Appl. Geophys.*, 116, 790-806.

Dieterich, J. H. (1979). Modelling of rock friction 1. Experimental results and constitutive equations, *J. Geophys. Res.*, 84, 2161-2168.

Dieterich, J. H. (1981). Constitutive properties of faults with simulated fault gouge, in

Mechanical Behavior of Crustal Rocks, *Geophys. Monogr. Ser.*, Vol. 24, 103-120.

Dorn, J. E. (1955). Some fundamental experiments on high temperature creep. *J. Mech. Phys. Solids*, 3, 85.

Duda, S. J. (1978). Physical significance of the earthquake magnitude-the present state of interpretation of the concept, *Tectonophysics*, 49, 119-130.

Eckmann, J. P. (1985). Ergodic theory of chaos and strange attractor, *Rev. Mod. Phys.*, 57, 617-656.

Engelder, J. T. (1974). Coefficients of friction for sandstone sliding on quartz gouge, *Advances in Rock Mech. Proc.* 3rd Congr. Int. Soc. Rock Mech., Denver, Washington D. C.: Nat. Acad. Sci., Vol II, part A, 499.

Farmer, J. D. (1982). Information dimension and the probabilistic structure of chaos, *Z. Naturforsch*, 37a, 1304-1325.

Farmer, J. D., E. Ott and J. A. Yorke (1983). The dimension of chaotic attractors, *Physica D*, 7, 153-180.

Feltham, P. (1966). *Deformation and Strength of Materials*, Butterworths: 72-75.

Froehling, H., J. P. Crutchfield, J. D. Farmer, N. H. Packard and R. S. Shaw (1981). On determining the dimension of chaotic flows, *Physica D*, 3, 605-617.

Fukunaga, K. (1972). *Introduction to Statistical Pattern Recognition*, Academic Press, New York.

Gabrielov, A. M., V. I. Keilis-Borok, T. A. Levshina and V. A. Shaposhnikov (1986). Block model of the dynamics of the lithosphere (in Russian), *Comput. Seismol.* 19, 168-178. (English translation, Allerton, New York.

Gardner, J. K. and L. Knopoff (1974). Is the sequence of earthquakes in southern California, with aftershocks removed Poissonian? *Bull. Seismol. Soc. Amer.*, 64, 1363-1367

Geng, N. (1986). The development of b-value simulated experiments and the beginning of b-value simulated experiments in China, *Acta Seismologica Sinica*, 8, 330-332.

Gong, C. (1982). *Thermodynamics and Statistical Physics*, Education Press, Beijing, (in Chinese).

Grassberger, P. and I. Procaccia (1983a). Measuring the strangeness of a strange attractor. *Physica D*, 9, 189-208.

Grassberger, P. and I. Procaccia (1983b). Estimation of the Kolmogorov entropy from a chaotic signal, *Phys. Rev. A*, 28, 2591-2593.

Green, A. G., C. P. Berry, C. Spencer, E. R. Kanasewich, S. Chiu, R. M. Clowes, C. J. Yorath, D. B. Stewart, J. D. Unger and W. H. Poole (1985). Recent seismic reflection studies in Canada, *AGU Geophysical Series*, Cornell Volume.

Green, A. G., R. M. Clowes, C. J. Yorath, C. Spencer, E. R. Kanasewich, M. T. Brandon and A. S. Brown (1986). Reflection mapping underplated oceanic

lithosphere and the subducting Juan de Fuca plate, *Nature*, 319, 210-213.

Griggs, D. T. and D. W. Baker (1969). *The Origin of Deep Focus Earthquakes, Properties of Matter Under Unusual Conditions*, Mark & Fernback, Interscience, New York.

Gutenberg, B. and C. F. Richter (1954). *Seismicity of the Earth*, Princeton Univ. Press, Princeton.

Gutenberg, B. and C. F. Richter (1956). Magnitude and energy of earthquakes, *Annali di Geofisica*, 9, 1-15.

Henon, M. (1983). Numerical exploration of Hamiltonian systems, in *Chaotic Behavior of Deterministic Systems*, eds. L Gerard, H. G. Robert and S. Raymond, North-Holland Publishing Company, Amsterdam.

Hirata, T., T. Satoh and K. Ito (1987). Fractal structure of spatial distribution of microfracturing in rock, *Geophys. J. R. Astr.* Soc., 90, 369-374.

Horowitz, F. (1989). A strange attractor underlying Parkfield seismicity, *EOS*, 70, 1359.

Huang, J. and D. L. Turcotte (1990). Are earthquakes an example of deterministic chaos?, *Geophys. Res. Lett.*, 17, 223-226.

Kagan Y. Y. and L. Knopoff (1987). Random stress and earthquake statistics: Time dependence, *Geophys. J. R. Astr. Soc.*, 88, 723-731.

Kagan, Y. Y. (1981a). Spatial distribution of earthquakes: The 3-point moment function, *Geophys. J. R. Astr. Soc.*, 67, 697-717.

Kagan, Y. Y. (1981b). Spatial distribution of earthquakes: The four-point moment function, *Geophys. J. R. Astr. Soc.*, 67, 719-733.

Kagan, Y. Y. (1988). Multiple expansions of extended sources of elastic deformation, *Geophys. J.*, 93, 101-114.

Kagan, Y. Y. and L. Knopoff (1978). Statistical study of occurrence of shallow earthquakes, *Geophys. J. R. Astr. Soc.*, 55, 67-86.

Kagan, Y. Y. and L. Knopoff (1980). Spatial distribution of earthquakes, the two-point correlation function, *Geophys. J. R. Astr. Soc.*, 62, 303-320.

Kanamori, H. and D. L. Anderson (1975). Theoretical basis of some empirical relations in seismology, *Bull. Seismol. Soc. Am.*, 65, 1073-1095.

Keen, C. E. and R. D. Hyndman (1979). Geophysical review of the continental of eastern and western Canada, *Can. J. Earth Sci.*, 16, 712-747.

Keilis Borok, V. I., V. G. Kosobokov, A. A. Solovjev and I. V. Kuznetzov (1986). *Curso Internacional Sobre Investigation Orientada a la Prediccion de Terremotos - Algoritmos, Programatica (Software) y Procesamiento de Datos*, Publication of Centro Regional de Sismologia Para America del Sur, Lima, Peru.

Keilis-Borok, V. I. (1990). The lithosphere of the earth as a nonlinear system with implications for earthquake prediction, *Rev. Geophys.*, 28, 19-34.

Keilis-Borok, V. I., L. Knopoff and I. M. Rotvain (1980a). Bursts of aftershocks, long-term precursors of strong earthquakes, *Nature*, 283, 256-263.

Keilis-Borok, V. I., L. Knopoff, I. M. Rotwain and C. R. Allen (1988). Intermediate-term prediction of occurrence times of strong earthquakes, *Nature*, 335, 690-694.

Keilis-Borok, V. I., L. Knopoff, I. M. Rovain and T. M. Sidorenko (1980b). Burst of seismicity as long-term precursors of strong earthquakes, *J. Geophys. Res.*, 85, 813-820.

Keilis-Borok, V. I., V. M. Podgayetskoya and A. G. Prozoroff (1971). On the local statistics of catalogs of earthquakes, *Computational Seismology*, 5, 55-79.

Kelleher, J. A. (1972). Rupture zones of large South American earthquakes and some predictions, *J. Geophys. Res.*, 77, 2087-2103.

Kelleher, J. A., L. Sykes and J. Oliver (1973). Possible criteria for predicting earthquake locations and their application to major plate boundaries of the Pacific and Caribbean, *J. Geophys. Res.*, 78, 2547-2585.

Knopoff, L. (1971). A stochastic model for the occurrence of main-sequence earthquakes, *Rev. Geophys. Space Phys.*, 9, 175-188.

Knopoff, L. and W. I. Newman (1983). Crack fusion as a model for repetitive seismicity, *Pure Appl. Geophys.*, 121, 495-510.

Kolmogorov, A. (1959). *Dolk. Akad. Nauk USSR*, 124, 754. English summary in MR 21, 2035.

Lamoreaux, R. (1982). *Cluster Patterns in Seismicity*. Ph.D. Thesis, University of Alberta, Edmonton.

Landau, L. D. and E. M. Lifshitz (1980). *Statistical Physics*, Pergamon Press, New York.

Li, Q. (1986). *Analysis of Seismic Instability of the Vancouver Island Lithoprobe Transect*, M.Sc. Thesis, University of Alberta, Edmonton.

Lorenze, E. N. (1963). Detecting nonperiodic flow, *J. Atmos. Sci.*, 20, 130-

Mahboobi, N. (1981). *Visco-elastic Instability and the Generation of Earthquakes*, Ph. D. Thesis, University of Newcastle upon Tyne, UK.

Mandelbrot, B. B. (1967). How long is the coast of Britain ? Statistical self-similarity and fractional dimension, *Science*, 156, 636-638.

Mandelbrot, B. B. (1983). *The Fractal Geometry of Nature*, Freeman, San Francisco.

Mandelbrot, B. B. and R. F. Voss (1983). Why is nature fractal and when should noises be scaling? *In Noise in Physical Systems and 1/f Noise*, edited by M. Savelli, G. Lecoy and J-P. Nougier, pp31-39, Elsevier Science Publishers B. V.

Meirovitch, L. (1970). *Method of Analytical Dynamics*, McGraw-Hill Book Co., New York.

Milne, W. G., G. C. Rogers, R. P. Riddihough, G. A. McMechan and R. D. Hyndman (1978). Seismicity of Western Canada, *Can. J. Earth Sci.*, 15, 1170-1193.

Mogi, K. (1985). *Earthquake Prediction*, Academic Press, New York.

Mogi, K. (1962a). Study of elastic shocks caused by the fracture of heterogeneous

materials and its relations to earthquake phenomena, *Bull. Earthq. Res. Inst.*, 40, 125-173.

Mogi, K. (1962b). Magnitude-frequency relation for elastic shocks accompanying fracture of various materials and some related problems in earthquakes, *Bull. Earthq. Res. Inst.*, 40, 831-853.

Mogi, K. (1979). Two kinds of gaps, *Pure and Appl. Geophys.*, 117, 1172-1186.

Mogi, K. (1981). Seismicity in western Japan and long-term earthquake forecasting, in *Earthquake Prediction*, An International Review. Maurice Ewing Ser., vol 4, edited by D. C. Simpson and P. G. Richards, pp43-51. AGU, Washington, D. C..

Newman, W. I. and L. Knopoff (1982). Crack fusion dynamics, a model for large earthquakes, *Geophys. Res. Lett.*, 9, 735-738.

Newman, W. I. and L. Knopoff (1983). A model for repetitive cycles of large earthquakes, *Geophys. Res. Lett.*, 10, 305-308.

Nyland, E. and Q. Li (1987). Analysis of seismic instability of the Vancouver Island lithoprobe transect, *Can. J. Earth Sci.*, 23,2057-2067.

Ogawa, M. (1987). Shear instability in a viscoelastic material as the cause of deep focus earthquakes, *J. Geophys. Res.*, 92, 13801-13810.

Ohnaka, M. (1973). Experimental studies of stick-slip and the application to earthquake source mechanism, *J. Phys. Earth*, 21, 285-303.

Osborne A. R. and A. Provenzale (1989). Finite correlation dimension for stochastic systems with power-law spectra, *Physica D*, 35, 357-381.

Packard, N. H., J. P. Crutchfield, J. D. Farmer and R. S. Shaw (1980). Geometry from a time series, *Phys. Rev. Lett.*, 45, 712-716.

Pathria, R. K. (1972). *Statistical Mechanics*, Pergamon Press, New York.

Rice, J. R. (1983). Constitutive relations for fault slip and earthquake instabilities, *Pure. Appl. Geophys.*, 121, 443-475.

Riddihough, R. P. (1979). Gravity and structure of an active margin __ British Columbia and Washington, *Can. J. Earth Sci.*, 16, 350-363.

Roux, J. C., R. H. Simoyi and H. L. Swinny (1983). Observation of a strange attractor, *Physica D*, 8, 257-266.

Rudnichi, J. W. (1988). Physical models of earthquake instability and precursory processes, *Pure. Appl. Geophys.*, 126, 531-554.

Ruelle, D. (1990) Deterministic chaos: the science and the fiction, *Proc. R. Soc. Land. A.*, 427, 241-248.

Ruina, A. L. (1983). Slip instability and state variable friction laws, *J. Geophys. Res.*, 88, 10359-10370.

Scholz, C. H. (1968). The frequency-magnitude relation of microfracturing in rock and its relation to earthquakes, *Bull. Seismol. Soc. Am.*, 58, 399-415.

Scholz, C. H. (1988). The critical slip distance for seismic faulting, *Nature*, 336, 761-763.

Scholz, C. H. and C. A. Aviles (1985). in *Earthquake Source Mechanics*, eds. S. Das, J. Boatwright and C. Sholz, 147-155, AGU Washington DC.

Scholz, C. H., P. Molnar and T. Johnson (1972). Detailed studies of frictional sliding of granite and implications for earthquake mechanism, *J. Geophys. Res.*, 77, 6392-6406.

Schuster, H. G. (1988). *Deterministic Chaos: an Introduction*, second revised edition, VCH Publishers, New York.

Shannon, C. E. and W. Weaver (1962). *The Mathematical Theory of Communication*, University of Illinois Press.

Shaw, R. (1985). *The Dripping Faucet as a Model Chaotic System*, The Science Frontier Express Series.

Shi, G. H. (1988). *Discontinuous Deformation Analysis, a New Numerical Model for the Statics and Dynamics of Block System*, Ph. D. thesis, University of California, Berkeley.

Shimazaki, K. and T. Nakata (1980). Time-predictable recurrence model for large earthquakes, *Geophys. Res. Lett.*, 7, No. 4, 279-282

Sieh, K. E. (1978a). Pre-historic large earthquakes produced by slip on the San Andreas fault at Pallet Creek, California, *J. Geophys. Res.*, 83, 3907-3939.

Sieh, K. E. (1978b). Slip along the San Andreas fault associated with the great 1857 earthquakes, *Bull. Seimsol. Soc. Am.*, 68, 1421-1448.

Sieh, K. E. (1981). A review of geological evidence for recurrence times of large earthquakes, in *Earthquake Prediction, An International Review*, Maurice Ewing Ser., vol 4, edited by D. C. Simpson and P. G. Richards, 181-207. AGU, Washington, D. C.

Sinai, Ya. G. (1959). *Dolk. Akad. Nauk USSR*, 124, 768. English summary in MR 21, 2036.

Smally, R. F. and D. L. Turcotte (1985). A renormalization group approach to the stick-slip behavior of faults, *J. Geophys. Res.*, 90, 1894-1900.

Suyehiro, S., T. Asada and M. Ohtake (1964). Foreshocks and aftershocks accompanying a perceptible earthquake in central Japan -- on a peculiar nature of foreshocks, *Pap. Neteoral. Geophys.*, 15, 71-78.

Sykes, L. R., and R. C. Quittmeyer (1981). Repeat times of great earthquakes along simple plate boundaries, in *Earthquake Prediction, An International Review*, Maurice Ewing Ser., vol 4, edited by D. C. Simpson and P. G. Richards, 217-247. AGU, Washington, D. C.

Takens, F. (1981). Detecting strange attractors in turbulence, *Lecture Notes in Mathematics*, 898, Springer, Heidelberg-New York.

Thatcher, W. (1984). The earthquake deformation cycle, recurrence, and the time-predictable model, *J. Geophys. Res.*, 89, 5674-5680.

Thompson, J. M. T. and H. B. Stewart (1986). Nonlinear Dynamics and Chaos, John Wiley and Sons, New York.

Turcotte, D. L. (1986). Fractals and fragmentation, *J. Geophys. Res.*, 91, 1921-1926.

Van Bueren, H. G. (1961). *Imperfection in crystals*, North Holland Publishing Co., Amsterdam, 233-235.

Vautard, R. and M. Ghil (1989). Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, *Physica D*, 35, 395-424.

Voss, R. F. (1989). Random fractals: self-affinity in noise, music, mountains, and clouds, *Physica D*, 38, 362-371.

Wallace, R. E. (1977). Profiles and ages of young fault scarps, north-central Nevada, *Geol. Soc. Am. Bull.*, 88, 1267-1281.

Week, J. and T. Tullis (1985). Frictional behavior of dolomite: A variation in constitutive behavior, *J. Geophys. Res.*, 90, 7821-7826.

Weertman, J. (1970). The creep strength of the earth's mantle, *Rev. Geophys. Space Phys.*, 8, 145-168.

Weertman, J. and J. R. Weertman (1975). High temperature creep of rock and mantle viscosity, *Annual Rev. Earth and Planet. Sci.*, 3, 293-315.

Weissman, M. B. (1988). 1/f noise and other slow, nonexponential kinetics in condensed matter, *Rev. Modern Phys.*, 60, No. 2, 537-571

Wiggins, S. (1988). *Global Bifurcation of Chaos*, Springer-Verlag, New York.

Xu, G. and H. Lan (1982). *Principle of Seismology*, Scientific Press, Beijing, (in Chinese).

Yorath, R. M., R. M. Clowes, A. S. Brown, M. T. Brandon, N. W. D. Nassey, A. G. Green, C. Spencer, E. R. Kanasewich and R. D. Hydnman (1984). Lithoprobe-Phase 1: South Vancouver Island: Preliminary analysis of reflection seismic profiles and surface geological studies, *Geological Survey of Canada Paper*, 85-1A, 543-554.

Yu, M. and F. Liu (1983). A theory of 1/f noise in metals, *Chinese Physics*, 3, No. 4, 863-876.

Appendix A

The finite element procedure is based on the well known principle of minimum total potential energy. Displacements are taken as the primary unknowns, therefore the appropriate functional of potential energy Π_p is

$$\Pi_p = \frac{1}{2}\sum\{d\}^T[k]\{d\} - \sum\{d\}^T\{r\} - \{D\}^T\{P\}$$

$$\{r\} = \int_V [B]^T[E]\{\varepsilon_0\}dv - \int_V [B]^T\{\sigma_0\}dv + \int_V [N]^T\{F\}dv + \int_S [N]^T\{\Phi\}dS$$

where $\{d\}$ is a nodal displacement vector. $\{D\}$ and $\{P\}$ are the structure vectors of nodal displacements and nodal forces. $\{\sigma_0\}$ and $\{\varepsilon_0\}$ are initial strain and stress vectors. $\{F\}$ is a body force vector and $\{\Phi\}$ is a surface traction vector. [N] is a shape function or interpolation function matrix. [B] is a strain-displacement matrix. [E] is a material matrix.

The infinitely thin fault element is described as a bilinear rectangular element subject to the limit $c \to 0$ (Figure A.1a). The shape function for a bilinear element is given as

$$[N] = \begin{bmatrix} N_1 & 0 & N_2 & 0 & N_3 & 0 & N_4 & 0 \\ 0 & N_1 & 0 & N_2 & 0 & N_3 & 0 & N_4 \end{bmatrix}$$

$$N_1 = \frac{1}{4bc}(b - x)(c - y)$$

$$N_2 = \frac{1}{4bc}(b + x)(c - y)$$

$$N_3 = \frac{1}{4bc}(b + x)(c + y)$$

$$N_4 = \frac{1}{4bc}(b - x)(c + y)$$

The strain-displacement matrix [B] is

$$[B] = \frac{1}{4bc}\begin{bmatrix} -(c-y) & 0 & (c-y) & 0 & (c+y) & 0 & -(c+y) & 0 \\ 0 & -(b-x) & 0 & -(b+x) & 0 & (b+x) & 0 & (b-x) \\ -(b-x) & -(c-y) & -(b+x) & (c-y) & (b+x) & (c+y) & (b-x) & -(c+y) \end{bmatrix}$$

For the infinitely thin element, we define the elastic constant as $\lim_{c\to 0} \frac{Y}{c} = Y_1$, which has to be finite in order to obtain the finite displacement across the thin fault element. In the case of plane strain, the stiffness matrix [k] is

$$[k] = \frac{bY_1}{12(1+\nu)} \begin{bmatrix} 2 & 0 & 1 & 0 & -1 & 0 & -2 & 0 \\ 0 & 4\lambda & 0 & 2\lambda & 0 & -2\lambda & 0 & -4\lambda \\ 1 & 0 & 2 & 0 & -2 & 0 & -1 & 0 \\ 0 & 2\lambda & 0 & 4\lambda & 0 & -4\lambda & 0 & -2\lambda \\ -1 & 0 & -2 & 0 & 2 & 0 & 1 & 0 \\ 0 & -2\lambda & 0 & -4\lambda & 0 & 4\lambda & 0 & 2\lambda \\ -2 & 0 & -1 & 0 & 1 & 0 & 2 & 0 \\ 0 & -4\lambda & 0 & -2\lambda & 0 & 2\lambda & 0 & 4\lambda \end{bmatrix}$$

where $\lambda = \frac{1-\nu}{1-2\nu}$, ν is Poisson ratio.

The initial strain $\{\varepsilon_0\}$ in above general formula can be considered as the strain due to inelastic deformation. We assume that inelastic deformation causes shear strain only, that is $\{\varepsilon_0\}=\{0, 0, \varepsilon_{12}\}^T$. ε_{12} equals $\Delta/2c$ by the definition of strain, where Δ is the inelastic displacement (Figure A.1b). The contribution of the inelastic deformation to {r} is

$$\int_V [B]^T[E]\{\varepsilon_0\}dv = \frac{bY_1\Delta}{4(1+\nu)} \{-1,0,-1,0,1,0,1,0\}^T$$

It forms elements in columns of the matrix A_2 in (1.9).

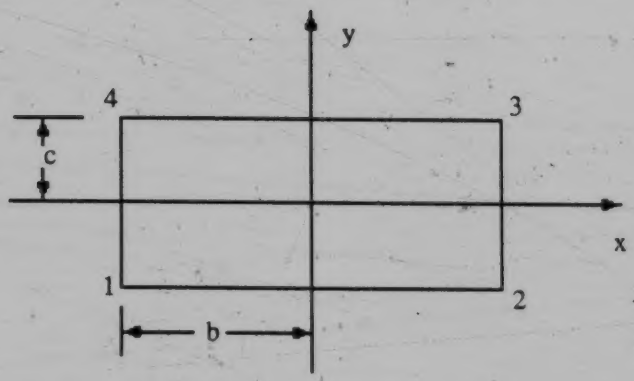
The average shear stress in an element is given by

$$\bar{\tau} = \frac{1}{4bc}\int_{-b}^{b}\int_{-c}^{c}[E]\left([B]\{d\} - \frac{1}{2c}\begin{Bmatrix} 0 \\ 0 \\ \Delta \end{Bmatrix}\right)dxdy$$

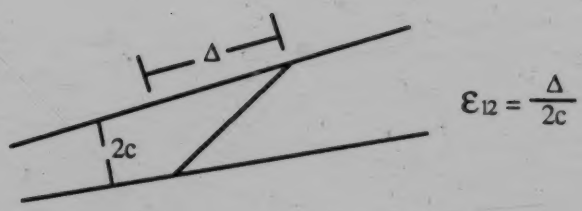
$$= \frac{Y_1}{4(1+\nu)}\left\{(-\frac{1}{2}, 0, -\frac{1}{2}, 0, \frac{1}{2}, 0, \frac{1}{2}, 0)\{d\} - \Delta\right\}$$

Comparing with the expression $\tau = \mathbf{M}_1\mathbf{d} + \mathbf{M}_2\Delta$ in the chapter one, we know that the elements in rows of $\mathbf{M}_1$ is formed by $Y_1/4(1+\nu)(-1/2, 0, -1/2, 0, 1/2, 0, 1/2, 0)$, and $\mathbf{M}_2$ is

a diagonal matrix with elements $Y_1/4(1+\nu)$. The matrix A_1 can be evaluated by following the standard procedure.



(a)



(b)

Figure A.1 (a) shows the definition of a binear element. (b)shows the difinition of shear strain.

Appendix B

i) The error in the correlation function

We define a random variable η_n as

$$\eta_n = \frac{\sum_{i,j=1}^{n} \xi_{ij}}{n^2} \qquad \text{where } \xi_{ij} = H(\varepsilon - |X_i - X_j|)$$

$H(\varepsilon - |X_i - X_j|)$ is the Heaviside step function, it have a value of 1 if the distance between x_i and x_j is less than ε, and 0 for the distance larger than ε. In the limit where n tends to infinity, η_n tends to the correlation integral $C(\varepsilon)$, which is the probability of two points separate within the distance ε. For the finite n, η_n is a random variable since it depends on the result of the particular n investigation. The mean of η_n is the probability.

Assumes ξ_{ij} has a probability of p to have the value of 1, and (1-p) for 0, and ξ_{ij} is independent of ξ_{lm} when i and j do not equal l and m. Since the mean of a random variable is the sum of all possible values multiplied by their respective probabilities, so we have

$$E(\xi_{ij}) = p$$
$$\sigma^2(\xi_{ij}) = E\{(\xi_{ij} - p)^2\} = (1 - p)^2 p + p^2(1 - p) = p(1 - p)$$

For η_n we have

$$E(\eta_n) = p = C(l)$$
$$\sigma^2(\eta_n) = \frac{1}{n^2} p(1 - p) < \frac{1}{n^2}$$

ii) The upper limit for the correlation dimension estimated from a finite length data series

Consider the slope in the Grassberger-Procaccia algorithm is measured as

$$\text{slope} = \frac{\log_a N(r'') - \log_a N(r')}{\log_a r'' - \log_a r'}$$

where $r' < r''$. Since $N(r') \geq 1$ and $N(r'') \leq 1/2N(N-1) \leq N^2$, we have following approximation,

$$\log_a N(r'') - \log_a N(r') \geq \log_a N^2$$

If we take $r'' \geq ar'$, we then have

$$\log_a r'' - \log_a r' \geq \log_a a$$

and therefore

$$\text{slope} \leq 2\log_a N$$

One may choose small base a to satisfy above relation, but it seams unreasonable to measure a slope over a very small length.

Appendix C

When a system depends on parameters E, a^1, a^2, ... a^s, the most probable distribution can be obtained by maximizing the entropy with the following constraints,

$$\sum_i P_i = 1$$

$$\sum_i P_i E_i = \overline{E}$$

$$\sum_i P_i a_i^m = \overline{a_i^m}$$

Introducing Lagarangian multiplies α, β, γ_1, ... γ_s, and maximizing the entropy, we obtain the distribution,

$$P_i = e^{-\alpha-\beta E_i+\sum_m \gamma_m a_i^m}$$

The entropy can be expressed as

$$S = \alpha + \beta\overline{E} - \sum_m \gamma_m \overline{a^m}$$

The change of S is

$$dS = \beta d\overline{E} - \sum_m \gamma_m d\overline{a^m}$$

Since the most probable distribution requires $dS = 0$, therefore

$$\gamma_m = \beta\left(\frac{\partial\overline{E}}{\partial\overline{a^m}}\right)_S = \beta\lambda_m$$

Replace γ_m by $\beta\lambda_m$, we have

$$P_i = e^{-\alpha-\beta\left(E_i-\sum_m \lambda_m a_i^m\right)}$$

$$\frac{1}{\beta} dS = d\overline{E} - \lambda_m d\overline{a^m}$$